STIC Database Tracking Number: 175929

TO: Manjunath N Rao
Location: REM-2A01/2C70
Art Unit: 1652
Tuesday, January 17, 2006
Case Serial Number: 10/618976

From: Kristine Hensle
Location: Biotech-Chem Library
REM-1B69
Phone: (571) 272-4161

Kristine. Hensle@uspto.gov

Search Notes

Examiner Rao,

See attached results.

If you have any questions about this search feel free to contact me at any time.

Thank you for using STIC search services!

Kristine Hensle
Librarian
STIC Biotech/Chem Library
(571)272-4161

175929

STIC-Biotech/ChemLib

From: Rao, Manjunath N.
Sent: Saturday, January 07, 2006 1:21 PM
To: STIC-Biotech/ChemLib
Subject: Sequence search request for 10619359

From: **Manjunath N. Rao**
      Art Unit 1652, Room 2A01
      Mail Box in Room 2C70
      Phone: 272-0939

Date: 1-7-06

Please search the following as soon as possible for application with serial number
10/618976

1. **SEQ ID NO: 1 against all <u>commercial nucleic acid sequence databases, issued</u> <u>patents/published applications nucleic acid sequence database</u> and <u>pending</u> <u>application nucleic acid sequence database.</u> Please provide a <u>print</u> of results**

2. **SEQ ID NO: 2 against all <u>commercial amino acid sequence databases, issued</u> <u>patents/published applications amino acid sequence database</u> and <u>pending</u> <u>application amino acid sequence database.</u> Please provide a <u>print</u> of results.**

If you have any questions please call me at the above phone number.

Thanks

STIC SEARCH RESULTS FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact:*

Mary Hale, Information Branch Supervisor
Remsen Bldg. 01 D86
571-272-2507

Voluntary Results Feedback Form

➢ *I am an examiner in Workgroup:* [＿＿＿]   *Example: 1610*

➢ *Relevant prior art* **found,** *search results used as follows:*

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)·
- ☐ Non-Patent Literature
  (journal articles, conference proceedings, new product announcements etc.)

➢ *Relevant prior art* **not found:**

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

sira
Search and Information

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2006, 10:11:00 ; Search time 7422 Seconds
(without alignments)
11511.144 Million cell updates/sec

Title: US-10-618-976-1
Perfect score: 1503
Sequence: 1 atgcaactccaacgcggctt.........ccgaggccagccagccgcgct 1503

Scoring table: IDENTITY_NUC
              Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*
              1: gb_ba:*
              2: gb_in:*
              3: gb_env:*
              4: gb_om:*
              5: gb_ov:*
              6: gb_pat:*
              7: gb_ph:*
              8: gb_pr:*
              9: gb_ro:*
              10: gb_sts:*
              11: gb_sy:*
              12: gb_un:*
              13: gb_vi:*
              14: gb_htg:*
              15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1503	100.0	1503	6	AR393878	AR393878 Sequence
2	647.4	43.1	1503	6	AY459352	AY459352 Thermus a
3	638.4	42.5	1503	6	E30186	E30186 Thermophili
4	638.4	42.5	2085	1	AB016244	AB016244 Thermus a
c 5	638.4	42.5	273285	1	AE017304	AE017304 Thermus t
c 6	636.8	42.4	1500	6	BD179859	BD179859 Highly th
7	636.8	42.4	1503	6	AX801458	AX801458 Sequence
c 8	636.8	42.4	110000	1	AP008226_11	Continuation (12 o
c 9	636.8	42.4	110000	1	AP008226_12	Continuation (13 o
10	596.2	39.7	1503	1	AY459351	AY459351 Thermus B
11	492.8	32.8	10747	1	AE002007	AE002007 Deinococc
12	421.8	28.1	110000	1	AE017180_12	Continuation (13 o
c 13	415.6	27.7	110000	1	BA000045_35	Continuation (36 o
c 14	403.2	26.8	110000	1	AP008231_05	Continuation (6 of
c 15	329.2	21.9	110000	1	BA000022_02	Continuation (3 of
16	326.4	21.7	307050	1	BX294140	BX294140 Pirellula
17	320.2	21.3	110000	1	BA000039_07	Continuation (8 of
18	314.8	20.9	110000	1	BA000019_46	Continuation (47 o
19	303.8	20.2	2204	15	AF307842	AF307842 Chlamydom
c 20	301.2	20.0	110000	1	AE017282_15	Continuation (16 o
c 21	293.8	19.5	110000	1	AY596297_17	Continuation (18 o
22	289	19.2	1518	6	AR480663	AR480663 Sequence
23	287.4	19.1	1515	6	AX571190	AX571190 Sequence
24	287.4	19.1	5910	6	CQ789074	CQ789074 Sequence
25	287.4	19.1	5910	6	AR218933	AR218933 Sequence
26	287.4	19.1	5910	6	BD003845	BD003845 Polynucle
c 27	287.4	19.1	11298	1	AE007499	AE007499 Streptoco
28	287.4	19.1	349980	6	AX571766	AX571766 Sequence
29	285.8	19.0	3569	1	STRMALMXP	STRMALMXP S.pneumonia
c 30	285.8	18.9	10864	1	AE008556	AE008556 Streptoco
c 31	284.4	18.9	300349	1	AE017319	AE017319 Desulfovi
32	284.2	18.9	137952	14	SPNEU1909	AL449931 Streptoco
33	265.6	17.7	1377	1	AR586217	AR586217 Sequence
c 34	211	14.0	110000	1	BA000016_26	Continuation (27 o
c 35	209.8	14.0	110000	14	CR954200_07	Continuation (8 of
c 36	209.8	14.0	110000	7	CR954202_06	Continuation (7 of
37	208.2	13.9	1806	15	AY570710	AY570710 Ostreococc
38	200.2	13.3	1731	15	AY081744	AY081744 Arabidops
39	200.2	13.2	1948	15	AY037231	AY037231 Arabidops
c 40	197.8	13.2	110000	1	CP000024_09	Continuation (10 o
c 41	196.2	13.1	110000	1	CP000023_09	Continuation (10 o
42	196	13.1	12278	1	AE009809	AE009809 Pyrobacul
43	184.8	12.3	9911	1	AE010052	AE010052 Streptoco
c 44	184.8	12.3	110000	1	CP000003_10	Continuation (11 o
45	181.4	12.1	2109	15	STDPEPA	X68664 Solanum tub

ALIGNMENTS

RESULT 1
AR393878
LOCUS   AR393878   1503 bp   DNA   linear   PAT 18-DEC-2003
DEFINITION Sequence 1 from patent US 6617143.
ACCESSION   AR393878
VERSION   AR393878.1 GI:40120997
KEYWORDS   .
SOURCE   Unknown.
ORGANISM   Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 1503)
AUTHORS   Fukuyama,S.
TITLE   Polypeptides having glucanotransferase activity and nucleic acids
            encoding same
JOURNAL   Patent: US 6617143-A 1 09-SEP-2003;
            Novozymes A/S; Bagsvaerd;
            DKX;
FEATURES   Location/Qualifiers
source   1..1503
            /organism="unknown"
            /mol_type="genomic DNA"

ORIGIN

Query Match   100.0%; Score 1503; DB 6; Length 1503;
Best Local Similarity   100.0%; Pred. No. 3e-243;
Matches 1503; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 ATGCAACTCCAACGCGCTTTTGGAATTTGCTCACCCCACCAGTTTCCGGGTCGCTCGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGCAACTCCAACGCGCTTTTGGAATTTGCTCACCCCACCAGTTTCCGGGTCGCTCGG 60

Qy   61 GGGATTGGGGCTCTGGGCCCGAGGCCGCCGAGCGGTTTTGGACTGGCTGGGCGATGCGGGA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGGATTGGGGCTCTGGGCCCGAGGCCGCCGAGCGGTTTTGGACTGGCTGGGCGATGCGGGA 120

Qy   121 GCCGCTGGTGGCAGGTCTTACCCGTCGGCCCTACCCAGTTACGGCGGACTCGCCGTACCAG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GCCCGCTGGTGGCAGGTCTTACCCGTCGGCCCTACCCAGTTACGGCGGACTCGCCGTACCAG 180

Qy   181 TCCTTCTCCGGCTTTTGCCCGTAACCCGTATTTGGTTGACCCGGACGTGATTGAATTGAAAAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   181 TCCTTCTCGGCTTTGCCGGTAACCCGTATTGTTGACCCCGAGATGCTGATTGAAAAA 240
Qy   241 GGCTGGCTGGAACAAAGCGAAGCGCCCCCGCCGTATCCGCCCAGCGCGTGGATTATGGC 300
Db   241 GGCTGGCTGGAACAAAGCGAAGCGCCCCCGCCGTATCCGCCCAGCGCGTGGATTATGGC 300
Qy   301 TGGCTTTACCAGACACCCGCTGGCCCCGTGTTCGCGGGGCTTTCGCGGGGTTTCGGCAAGG 360
Db   301 TGGCTTTACCAGACACCCGCTGGCCCCGTGTTCGCGGGGCTTTCGCGGGGTTTCGGCAAGG 360
Qy   361 GCTTCGGCGGCCAGGATAAGACCCGACTGGAAGCCCTTTATCGAGGCCGAGCGCTTCGGCTG 420
Db   361 GCTTCGGCGGCCAGGATAAGACCCGACTGGAAGCCCTTTATCGAGGCCGAGCGCTTCGGCTG 420
Qy   421 GAAGACTATGCGCTTCTTTATGCGCGACCCGGTTTGACGGCAAGCCCTGGAACCAG 480
Db   421 GAAGACTATGCGCTTCTTTATGCGCGACCCGGTTTGACGGCAAGCCCTGGAACGAG 480
Qy   481 TGGAGCCCCGAGCTGCGCGACCGGTGAACCGGCTGCCCTGCCCAGGGCCCGTGAGGAGCTG 540
Db   481 TGGAGCCCCGAGCTGCGCGACCGGTGAACCGGCTGCCCTGCCCAGGGCCCGTGAGGAGCTG 540
Qy   541 GCCGAGGAGGTGGCCGCTTTACGAGTGGGATTCAGTGCGCTTTTATCTGGAATGGGGCCAGG 600
Db   541 GCCGAGGAGGTGGCCGCTTTACGAGTGGGATTCAGTGCGCTTTTATCTGGAATGGGGCCAGG 600
Qy   601 ACCAAGGCCTATGCCAAGGGGATTCAGATTATCGGCGATATGCCCATCTTTGTG 660
Db   601 ACCAAGGCCTATGCCAAGGGGATTCAGATTATCGGCGATATGCCCATCTTTGTG 660
Qy   661 GCCTTCGATTCCTCAGATGTCTCGGGCCAACCCGCAGTACTTCTACCTCGAGGCCGATGGC 720
Db   661 GCCTTCGATTCCTCAGATGTCTCGGGCCAACCCGCAGTACTTCTACCTCGAGGCCGATGGC 720
Qy   721 AACCCCACGGTGGTGGCGGCGGTTCCGCGGGACTACTACTTCTCCGAAACCGGCCAGCTCTGG 780
Db   721 AACCCCACGGTGGTGGCGGCGGTTCCGCGGGACTACTACTTCTCCGAAACCGGCCAGCTCTGG 780
Qy   781 GGCAATCCGCTCTATCGCGTCGGGATGTGATGGAAAGGGACAACTTTGCCTTGGTGCATTGCC 840
Db   781 GGCAATCCGCTCTATCGCGTCGGGATGTGATGGAAAGGGACAACTTTGCCTTGGTGCATTGCC 840
Qy   841 CGCATAAGGCAGTCGCTCAAGCAGTGCCACCTGGTCGCATGCGACGTCCACCACTTCCGCGGGTTT 900
Db   841 CGCATAAGGCAGTCGCTCAAGCAGTGCCACCTGGTCGCATGCGACGTCCACCACTTCCGCGGGTTT 900
Qy   901 GAAGCCTACTGGGAGGTTCCGTTTGGCTGCGCGGTGCGGGTGGAGAGGGGCGCTGGGTCAA 960
Db   901 GAAGCCTACTGGGAGGTTCCGTTTGGCTGCGCGGTGCGGGTGGAGAGGGGCGCTGGGTCAA 960
Qy   961 GCCCCAGGGGAGAGAAGCTGTTTGCTGCGGTGCGGGTGCGGGCCCAACTGAGCGCATGGCCCCATCATT 1020
Db   961 GCCCCAGGGGAGAGAAGCTGTTTGCTGCGGTGCGGGTGCGGGCCCAACTGAGCGCATGGCCCCATCATT 1020
Qy 1021 GCCGAAGACCTGGGGGTGATCACCCCGGAGGTGGGAGGCTTTGCGCGATGGCTTCGGGTTC 1080
Db 1021 GCCGAAGACCTGGGGGTGATCACCCCGGAGGTGGGAGGCTTTGCGCGATGGCTTCGGGTTC 1080
Qy 1081 CCCGGCATGCGGTGCTTTTTGCAGTTTGCGGTTTTCCGGGTGGAGGACAACCCCACCACC 1140
Db 1081 CCCGGCATGCGGTGCTTTTTGCAGTTTGCGGTTTTCCGGGTGGAGGACAACCCCACCACC 1140
Qy 1141 AACTACCCCGCCACCGGACCGCCACCGGACCGGAACCCAGCGACACCACC 1200
Db 1141 AACTACCCCGCCACCGGACCGCCACCGGACCGGAACCCAGCGACACCACC 1200
Qy 1201 CTGGGATGGTTCCGCACCGGACCGAGCGGCCGCCTTCATGCGGGCGGCTTACCTGGCC 1260
Db 1201 CTGGGATGGTTCCGCACCGGACCGAGCGGCCGCCTTCATGCGGGCGGCTTACCTGGCC 1260
Qy 1261 CGCTATGGCATCCGTGTTGTTGTCGGAATACGAGGTGCGGGCGCTTTGATGAGCTGGCC 1320
Db 1261 CGCTATGGCATCCGTGTTGTTGTCGGAATACGAGGTGCGGGCGCTTTGATGAGCTGGCC 1320

Qy 1321 TTCAAAAGCCCGGCCAAGCTGGCTATTGTGCTTTGCAGGACGTGCTGGGGCTGGGCCCCC 1380
Db 1321 TTCAAAAGCCCGGCCAAGCTGGCTATTGTGCTTTGCAGGACGTGCTGGGGCTGGGCCCCC 1380
Qy 1381 GAGGCCCGCATGAACTTCCCCGACGGCTGGGACGGCTGGGGACAACTGGGCGGCTGGGCGTGGGCGTACGCCGAA 1440
Db 1381 GAGGCCCGCATGAACTTCCCCGACGGCTGGGACGGCTGGGGACAACTGGGCGGCTGGGCGTGGGCGTACGCCGAA 1440
Qy 1441 GGGACCTTCGAGCCCGGTCTGGCGCCGCTGGGGCCGGGCACTGCGGGCCCTGGGCCCGGCAGGCCAGCGC 1500
Db 1441 GGGACCTTCGAGCCCGGTCTGGCGCCGCTGGGGCCGGGCACTGCGGGCCCTGGGCCCGGCAGGCCAGCGC 1500
Qy 1501 GCT 1503
Db 1501 GCT 1503

RESULT 2
AY459352
LOCUS   AY459352   1503 bp   DNA   linear   BCT 30-NOV-2003
DEFINITION Thermus aquaticus 4-alpha-glucanotransferase gene, complete cds.
ACCESSION   AY459352
VERSION   AY459352.1 GI:38505480
KEYWORDS   .
SOURCE   Thermus aquaticus
  ORGANISM Thermus aquaticus
            Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
            Thermus.
REFERENCE   1 (bases 1 to 1503)
  AUTHORS   Park,K.-H. and Park,J.-H.
  TITLE   Direct Submission
  JOURNAL   Submitted (05-NOV-2003) Food Science & Technology, Seoul National
            University, San 56-1, Shillim-dong, Kwanak-gu, Seoul 151-742, Korea
FEATURES   Location/Qualifiers
     source   1..1503
                     /organism="Thermus aquaticus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:271"
     CDS   1..1503
                     /codon_start=1
                     /transl_table=11
                     /product="4-alpha-glucanotransferase"
                     /protein_id="AAR23242.1"
                     /db_xref="GI:38505481"
                     /translation="MELPRAFGLLLHPTSLPGPYGVGVLGLEARDFLRFLKGAGGRFW
                     QVLPLGPTGYGDSPYQAFSAFAGNPYLIDLRPLAEBKGYLVLKDPGFPQGRVDYGWLYA
                     WKWPALKAAYQGFPLEKAPRKEREDFPAFQEKEASWLKDYALFMALKAQHGGLPWNRWP
                     LPLRRREKAMKERAALBEVAFHAWTOWLFPFEAWKALKEEAEALGIQIGDMPIFV
                     AEDSAEVWAHPEWFHLDEEGRPLVVAGVPPDYFSETGQRWGNPLYRWDVLEREGFSFW
                     IARLAKALEFLHLVRVDHNFRGFEAYWEIPASCPTAVEGRWVKASGEKLFDRIQEVFGE
                     VPILABDLGVITPEVEPLRDRYGLPGMKVLQFAFDHGMENPFLPHNYPANGRVVYTG
                     THDNDTTLGWYRTASPHERAFLERYLADWGISFRQEBEVPWALMGLCLKSVARLAIYP
                     VQDVLALGSEARMNYPGRPSGNWAWRLLPGQLTQBHAARLLAMABATGRT"

ORIGIN
Query Match   43.1%;   Score 647.4; DB 1; Length 1503;
Best Local Similarity   65.6%;   Pred. No. 3.3e-99;
Matches   980; Conservative   0; Mismatches 506; Indels   9; Gaps   2;

Qy   1 ATGCAACTCCAACGCGGCTTTGGGGCGCTTTGGAATTTGCTCCACCCCACCAGTTTCTCCGGTCGCTGG   60
Db   1 ATGGAGCTTCCTAGGGCGTTTGGGCTGCTCCTCCATCCCACGAGCCTCCCGGGCCCCTAC   60
Qy   61 GGGATTGGGGCTTCTGGGCGGTCTCGGGGCGGGGTTTTGGACTGGCTGGCCATGCGGGGA 120
Db   61 GGCGTGGGGGTCCTGGGCCTCGAGGCCCGCGACTTCCTCCGCTTCCTAAAAGGGGCCGGG 120
Qy   121 GCCCGCGCTTCTGGCAGGTCTTACCGGTGCTGGGCGTGGGCCTGGGCCGTACCAGTACCAG 180
Db   121 GGGCGCTTCTGGCAGGTCCTCCCCCTGGGGCCGACCGGGTACGGGGACTCCCCTACCAG 180
Qy   181 TCCTTCTCGGCTTTGCCGGTAACCCGTATTGTTGCCGGTAACCCGTATTGGTTGACCCCGAGATGCTGATTGAAAAA 240

```
Db   1255  GCGGACTGGGGGATCAGCTTCCGCCAGGAGGAGGAGTGCCCTGGGCCCTGATGGGGCCTT 1314
Qy   1318  GCCTTCAAAAGCCCGGCCAAGCTGCTATTGTGCCTTGCCTTTGCAGGACGTGCTGCGGGTCGGCTCGGGGC 1377
Db   1315  TGTTTGAAGTCCGTGGCCCAGGCTGGGCCATTTACCCCGTGCGAGGACGGTTCTCGCCCTCGGGC 1374
Qy   1378  CCCGAGGCCCGATGGAACTTCCCCGACAGCGGCGTGGGGGGACAACTGGGCGTGGGCCTACGCC 1437
Db   1375  TCCGAAGCCCCGATGAATGAACTACCCGGCGCGCCCTTCGGGCAACTCGGGCTCGGGGCGGTCCTT 1434
Qy   1438  GAAGGCGACCTCGAGCCCGGTCTCGGCCCGGGGACTGCGGGCCCTGGCCGAGGCCA 1492
Db   1435  CCGGGGGCAGCTCACCCAGGAGCACGCGCCCGTCTCCTGGCCATGGGCGGAGGCCA 1489
```

```
Query Match      42.5%;  Score 638.4;  DB 6;  Length 1503;
Best Local Similarity   65.0%;  Pred. No. 1.1e-97;
Matches 977;  Conservative 0;  Mismatches 516;  Indels 9;  Gaps 2;

Qy     1  ATGCAACTCCAACGCGCTTTTGGAATTTGCTCACCCACCAGTTTCCGGTCGTCGTGG   60
Db     1  ATGGGAGTTCCCCGGCGCTTTCGGGTCGTCGCCGAGGCCGGTTTTGGACTGGCTGCCCGGCCTCCCCGGCCCCTAC   60
Qy    61  GGGATTGGGGCTTGGGCGCCGAGGCCGGTTTTGGACTGGCTGCGGCCGATGCGGGGA  120
Db    61  GGGCGTCGGGCGTCTGGCCGGCGGGGACTTCCTCCGCCTTCCTCAAGGAGGCGGGGG  120
Qy   121  GCCCGCTGGTGGCAGGTCTTACCGGTGGGCCCTACCAGTTACGGCGGACTCCGTACCAG  180
Db   121  ACCCTGGGGCTGGTACCGGCGGCCTTCCTGGAGCGGCGGTTACCTG  180
Qy   181  TCCTTCTCGCGTTTTGTTGTCCGGTAACCGTATTTGGTTGACCCCGAGTGCTGATTGAAAAA  240
```

```
Db  1255  GCGGACTGGGGATCACCTTCCGGGAAGAGAGGAGGTGCCTGGGCCCTGATGCACCTG  1314
               ||||||  |||  |||  ||||||||||  |  | ||||||||||| | |||||||||
Qy  1318  GCCTTCAAAAGCCCGGCCAAGCTGCTATTGTGCCTTTGCAGGACGTGCTGGGGCTCGGGC  1377
               ||  |||||||||  |  |||    ||  ||| |||||||||  ||||||||| |||
Db  1315  GGGATGAAGTCCGTGCCCGGCTCGCCGTCTCACCCGGTGCAGGAGTCTGGCCCTGGGC  1374
               ||||||||||  |  || ||||||||  || ||||  ||||||||  |||||||||
Qy  1378  CCCGAGGCCCGCATGAACTTCCCCGGACACGGCTGGGGGACAACTGGGCGTGGCTACGCC  1437
               ||||||||||||  ||  ||||||||  |||||||| ||||||||||||| ||||||
Db  1375  AGCGAGGCCCGGATGAACTACCCCGGGAAGGCCCTCGGGGAACTGGGCCTCGGGCTCCTC  1434
               |||||||||||||||  ||||||||||  ||| | ||||||||||||||||||||
Qy  1438  GAAGGCGACCTCGAGCCCGGTCTGGCCGCGGGACTGCGGGGCCCTGGCCAGCCAGCCAG  1497
               |||||||||||  ||| ||||||||| ||  ||||||||||||  ||| |||||||
Db  1435  CCGGGGGAGCTTTCCCGGAGCACGGGCAGGCTTAGGGCGGCCATGGCCGGAGCCACGGAA  1494
               |||  |  |||
Qy  1498  CG  1499
          ||
Db  1495  CG  1496
```

RESULT 4
```
LOCUS        AB016244                  2085 bp    DNA      linear    BCT  24-MAR-1999
DEFINITION   Thermus aquaticus gene for amylomaltase, complete cds.
ACCESSION    AB016244
VERSION      AB016244.1  GI:3721879
KEYWORDS     amylomaltase.
SOURCE       Thermus aquaticus
  ORGANISM   Thermus aquaticus
             Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
             Thermus.
REFERENCE    1  (sites)
  AUTHORS    Terada,Y., Fujii,K., Takaha,T. and Okada,S.
  TITLE      Thermus aquaticus ATCC 33923 amylomaltase gene cloning and
             expression and enzyme characterization: production of cycloamylose
  JOURNAL    Appl. Environ. Microbiol. 65 (3), 910-915 (1999)
   PUBMED    10049841
REFERENCE    2  (bases 1 to 2085)
  AUTHORS    Terada,Y., Fujii,K., Takaha,T. and Okada,S.
  TITLE      Direct Submission
  JOURNAL    Submitted (14-JUL-1998) Yoshinobu Terada, Ezaki Glico Co., Ltd.,
             Biochemical Research Laboratory; Utajima 4-6-5, Nishiyodogawa-ku,
             Osaka, Osaka 555-8502, Japan (E-mail:terada-yoshinobu@glico.co.jp,
             Tel:81-6-477-8425, Fax:81-6-477-8271)
FEATURES             Location/Qualifiers
     source          1..2085
                     /organism="Thermus aquaticus"
                     /mol_type="genomic DNA"
                     /strain="ATCC33923"
                     /db_xref="taxon:271"
     gene            514..2016
                     /gene="malQ"
     CDS             514..2016
                     /gene="malQ"
                     /codon_start=1
                     /transl_table=11
                     /product="amylomaltase"
                     /protein_id="BAA33728.1"
                     /db_xref="GI:3721880"
                     /translation="MELPRAFGLLLHPTSLPGPYGVGVGVLGREARDFLRFLKEAGGRYW
                     QVLPLGPTGYGDSPYQSFSAFAGNPYLIDLRPLAERGYVRLEDPFPQGRVDYGLLYA
                     WKWPALKGARPGFKEKASPBEREAFAAFPEREAWWLEDYALFMALKGAHGGLPWNRWP
                     LPLRKREBKALREAKSALABEVVAFHAPTQWLFPRQWGALKABEALGIRIIGDMPIFV
                     AEDSAEVWAHPENWFHLDEBEGRPTVVAGVVPPDYFSETGQRWGNPLYRWDVLEREGSFW
                     IRRLEKALELFHLVRIDHFRGFEAYWEIPASCPTAVEGRWVKAPGEKLFQKIQEVFGE
                     VPVLAEDLGVITPEVEALRDRFGLPGMKVLQFAPDDGMENPFLPHNYPAHGRVVVYTG
                     THDNDTTLGWYRTATPHEKAFMARYLADWGITFREEEEVPWALMHLGMKSVARLAVYP
                     VQDVLALGSEARMYPGRPSGNWAWRLLPGELSPEHGARLRAWABEATERL"
```

ORIGIN

```
Query Match          42.5%;    Score 638.4;    DB 1;    Length 2085;
Best Local Similarity    65.0%;    Pred. No. 1e-97;
Matches 977;    Mismatches 516;    Conservative 0;    Indels 9;    Gaps 2;
```

```
Qy  181  TCCTTCAGCGCCTTCGCGCGGAAACCCCTACCTCATAGACCTGAGGCCCCTGCGGAAAGG  240
              ||||||  ||||||  ||||||||||  |  ||||  | ||||  |||||||| |||||
Db  241  GGCTGGCTGGAACAAAGGGAAGCGCCCCCGCCTGCCGGCGTGGGATTATGGC  300
              |||||        |||||||  ||||||||||  ||||||||| |||||||
Qy  241  GGCTACGTG------CGCCTGGAGACCCCGCTTCCCCGAAGGCGGGTGGGACTACGGC  294
              |||||  ||||||||||  ||||||  ||||||||||| ||||||||||||||||
Db  301  TGGCTTACCAGACCCGCTGGCCCGTTGCGGCGGGCTTTCGGGGTTTCGGGCAAGG  360
              |||||||||||  ||||||||||||||  ||||||||||||||||||||
Qy  295  CTCCTCTACGCTGGAAGTGGCCCGCCCTGAAGGAGGCCCTTCCGGGGCTTCAAGGAAAAG  354
              |||||||||||||||||||||||||||||||  ||||||||||||||||
Db  361  GCTTCGGCCCAGGATAAGACCCGACTGGAAGCCCTTTATGGAGGGCCTCTGGCTG  420
              |||||||  |||||  |||||||||  ||||||||||||||||||||||
Qy  355  GCCTCCCGGAGGAGGCGGGAGGCCTTCGCCGCCCTTCGCCGCACGGGAGGAGGCCTGGTGGCTC  414
              ||||||||||||  |||||||  |||  |||||||||||  |||||||||||||
Db  421  GAAGACTATGCGCTCTTTATGCCCTCAAGACCCGGTTTGACGGGCAAGCCCTGGAACGAG  480
              |||||||||  |  ||  ||||||||||||||||||||||||||||||||
Qy  415  GAGGACTACGCCCTTCTTCATGCCCCTGAAGGGCCGCACGGGGGGCTTCCCTGGAACCGG  474
              |||||||||||||||||||||||||||||||||||||||||||
Db  481  TGGAGCCCCCAGCTGCGGGACCGTGAACCGGCTGCCCTGCCCAGGCCCCGTGAGGAGCTG  540
              |||||||||||||||||||||||||||||||||||||||
Qy  475  TGGCCCCCTTCCCCTGCGCAGGCGGAAGAGAAGAGGCCCTTAGGGAGGAGGGAAAAAGGCCGCCTG  534
              |||||||||||||||||||||||||||||||||||||||
Db  541  GCCGAGGAGTGGGCCCTTTACGAGTGGATTCAGTGGCCTTTTTATCGTGGAATGGGGCCAG  600
              |||||||||||||||||||||||||||||||||||
Qy  535  GCCGAGGAGTGGGCCTTCACCGCCGTCTCTTCTTCCGCCAGTGGGGGGGGGCC  594
              |||||||||||||||||||||||||||||||||||
Db  601  ACCAAGGCCTATGCCGAATCCAAGGGGATTCAGATTATCGGCGATAATGCCGATCCCATCTTTGTG  660
              |||||||||||||||||||||||||||||||||||
Qy  595  TTGAAGGCGGAGGCCGAGGCGTTGGGCATCCGGACATCGGGCGACATGCCCATCTTCGTG  654
              |||||||||||||||||||||||||||||||||||
Db  661  GCCTTCGATTCCTCAGATGTCTCGGGCCAACCGGCTACTTCTACTTCGAGGGCCCATGGC  720
              |||||||||||||||||||||||||||||||||||
Qy  655  GCCGAGGACTCCGCGCCCACCGAGTCGTTTCACCTGGACGAGGAGGAGGGC  714
              |||||||||||||||||||||||||||||||||||
Db  721  AACCCCACGGTGGTGGCGGCGTTCCGGCGGACTACTTCTCCGAAACCGGCCAGCTCTGG  780
              |||||||||||||||||||||||||||||||||||
Qy  715  CGCCCCCACGGTGGTGGCGGGGGTGCCCCCGACTACTTCTCCGGGAGACGGGCCAGCGCTGG  774
              |||||||||||||||||||||||||||||||||||
Db  781  GGCAATCCCTCTATCGTGGATGTGATGGAAAGGGACAACTTTGCCTGGTGCATTGCC  840
              |||||||||||||||||||||||||||||||||||
Qy  775  GGTAACCCCTTTACCGCTGGACGTTTTGAGCGGGGGAGGGGTTCTCCTTCTGGATCCGC  834
              |||||||||||||||||||||||||||||||||||
Db  841  CGCATAAGGCAGTCCTCAAGCAGTGCCACTGGTGCCATCGACCACTTCCGCGGGTTT  900
              |||||||||||||||||||||||||||||||||||
Qy  835  CGTCTGGAGAAGGCCCTGGAGCTCTTCCACCTGGTGCGCATAGACCACTTCCGCGGGCTTT  894
              |||||||||||||||||||||||||||||||||||
Db  901  GAAGCCTACTGGGAGGTTCCGTTTGGCCGGCAAATGCTGTGAGGGGGGCTGGGTCAA  960
              |||||||||||||||||||||||||||||||||||
Qy  895  GAGGCCTACTGGGAGATCCCCGCAAGCTGCCCCACGCCGGCGGGGCCAACTGGGTCAAG  954
              |||||||||||||||||||||||||||||||||||
Db  961  GCCCCCAGGGGGAGAGCTGTTGCTGCGTGGTGCGGGGCCCATCATT  1020
              |||||||||||||||||||||||||||||||||||
Qy  955  GCCCCCGGGGGGAGAAGCTCTTCCAGAAGATCCAGGAGGTCTTCGGCGAGGTCTTCCTC  1014
              |||||||||||||||||||||||||||||||||||
Db  1021  GCCGAAGACCTGGGGGTGATCACCCCCGAGTGGGAGGCCTTTGCGGGATGGCTTCCGGGTTC  1080
              |||||||||||||||||||||||||||||||||||
Qy  1015  GCCGAGGACCTGGGGGTCATCACCCCCGAGTGGGAGGCCGACCCGTTCGGGGGCCCTT  1074
              |||||||||||||||||||||||||||||||||||
Db  1081  CCCGGCATGAAGATTTTGCAGTTTGCTTT---TCCGGTGAGGACAACGGCCTTTTGCCC  1137
              |||||||||||||||||||||||||||||||||||
Qy  1075  CCCGGGATGAGAGTCCTGCAGTTCGCCTTTGACGCGGGATGGAAAACCCCTTCCTCCCCC  1134
              |||||||||||||||||||||||||||||||||||
Db  1138  CACAACTACCCCGGCACCCCGCAATGTGGTGTGTACAGCGGCAATGGCGGAACCCACGACACC  1197
              |||||||||||||||||||||||||||||||||||
Qy  1135  CACAACTACCCTGCCAGCCGCCGGGTGGTCTACACCGGACCACCACGACACC  1194
              |||||||||||||||||||||||||||||||||||
Db  1198  ACCCTGGGATGGTTCCGCACCGGCCTTCATGCCGGCCTTACCTG  1257
              |||||||||||||||||||||||||||||||||||
Qy  1195  ACCCTGGGTCGGTACCGCACCGGCCTTCATGCGGCGGTACCTG  1254
              |||||||||||||||||||||||||||||||||||
Db  1258  GCCGCTATGGCATCCGTTGTTTGTCGGAATACGAGGTGCGCGGGCGCTTTGATCGAGCTG  1317
```

Qy   1 ATGCAACTCCAAGCGGCGTTTTGGAATTTGCTCCACCCCACCAGTTTCCGGGTCGCTGG 60
           |||| || ||| ||||||| || |||||| || ||||||||||| || || |||
Db 514 ATGGAGCTTCCCCGGGCTTCGGGTCTGCTTCCCACCCCACGAGCCTCCCCGGCCCCTAC 573

Qy   61 GGGATTGGGCGGCTCTGGGCCGCGAGGCCGAGCGGTTTTGGACTGGCTGGCTGATGCGGGA 120
           || |||| | |||| ||| |||||||||| || ||||||||||| || ||||||||||
Db 574 GGCGTCGGCGTCTGGGCCGGCGGGAGGCCCGGGGACTTCCTCCGGTTCCTCCTCAAGGAGG 633

Qy 121 GCCCGCTGGTGGCAGGTCTTACCGCTGGGCCTACCAGTTACGGGCGACTGCCGTACCAG 180
           || || ||||| |||| ||| || |||||||| ||||| || ||| |||| |||||||
Db 634 GGGCGGTACTGGCAGGTGCTCTCCCCACGGGCTATGCGGCGACTCCCCCTACCAG 693

Qy 181 TCCTTCTCGGCTTTTGCCGGTAACCCGTATTTGGTTGACCCGAGATGCTGATTGAAAA 240
           |||||| ||||||| || |||||| ||| || |||||| |||||||||| || |||
Db 694 TCCTTCAGCGGCTTCGCCCGGGAAACCCCTCATGAGACCTCATAGACCTGAGCCCCTCGGCGAAAGG 753

Qy 241 GGCTGGCTGAACAAAGCGAAGCGCCCCGCCTATCCGACCCAGCGCGTGGATTATGGC 300
                    ||| |||||||||||| || |||||| |||||||||||||||
Db 754 GGCTACGTG------CGCCTGGAGGACCCCGGCTTCCCCCAAGGCCCGGGTGACTACGGC 807

Qy 301 TGGCTTTACCAGACCCGCTGCTGCCCCTGTTGCGGCGGTTTCGGCGGGTTTCGGCGCAAGG 360
            |||||| || |||||| |||||||| || |||||||||| || ||||||| |||||||
Db 808 CTCCTCTACGCGGCTGGAAGTGCCCGCGCCTTGAAGGAGGGCCTTCCCGGGGCTTCAAGGAAAAG 867

Qy 361 GCTTCGCCCAGGATAAGACCCGACTGGAAGCCTTTATCGAGGCGCTTCTGCTG 420
           || |||||| || |||||| || ||| || |||||| |||| |||| ||| ||
Db 868 GCCCCCCCGGAGGAGCGGGGAGGCGGGGAGGCCTTCGCCGGCCTTCGCCGGAGGGGGCTC 927

Qy 421 GAAGACTATGCGCTCTTTATGGCCCTCAAGACCCGGTTTGACGGCAAGCCCTGGAACGAG 480
            ||||| |||||||| ||| ||||| | |||| |||||||||| ||| |||||||| |
Db 928 GAGGACTACGCCCCTCTTCATGGCCCTGAAGGGGGCCGCACCGGGGGCTTCCCCTGGAACCCGG 987

Qy 481 TGGAGCCCCGAGCTGCGGCCTGCGACCGGTGAACCGTGCCCTGCCAGGGCCCGTGAGGAGCTG 540
           ||||| ||||| |||||| ||||||||||| || |||| || |||| |||||| ||||||
Db 988 TGGCCCCTTCCCCTGCGCGGAAGCGGGAAGAGAGCCCTTAGGGAGGCGAAAAGCGCCTTG 1047

Qy 541 GCCGGAGGTGGGCCCTTTACGAGTTGGATTCAGTGCGTTTTTTTATCGGAATGGGGCCAGC 600
           || ||||||||||||||| ||| ||| | |||||||||||| || |||| ||||| || |
Db 1048 GCCGGAGGAGGTGGGCCTTGCCACCGGGCTTCCACCGTGGGGGGGGCC 1107

Qy 601 ACCAAGGCCTATGCGCAATCCAAGGGGATTCAGATTATCGGCGATATGCGGCCATCTTGTG 660
           |||||||| | |||| |||| || |||||| ||| ||| |||| ||| |||| ||| || |
Db 1108 TTGAAGGCGAGGAGGGCGGTTGGGCATCCGATGGGGGGGATCATCGGCGATCAGCATCTTCGTG 1167

Qy 661 GCCTTCGATTCCTCCTAGAGTCTGGGGCCAACCGCAGTACTTCTACCTCGAGGCCGATGGC 720
           || |||||||| ||| || |||| ||||| || |||||| ||| || |||| ||||||||
Db 1168 GCCGAGGACTTCGCCGCGGAGGTCGGGGCCACCCCCGAGTGGTTCCACCTGGACGAGGAGGGC 1227

Qy 721 AACCCCACGGTGGTGCGGCGGGTTCGCGGGACTACTTCTTCCGGAAACCGGCCAGCTCTGG 780
           |||| |||||| ||||| || |||||| ||||| || || |||||||| ||||||||||||
Db 1228 CGCCCCCACGGTGGTGCGGTGCGGGGGTGCCCCACCTACCCGGGAGACGGGGCCCAGCGCTGG 1287

Qy 781 GCAATCCGCTCTATCGCTGGGATGTGATGCGAAAAGGGACAACTTGCCTGGTGCATTGCC 840
           || ||||||| || || ||||| |||| ||||| ||| || ||||| || || ||||||
Db 1288 GGTACCCCCTTTACCGCCTGGACGTTTGGAGCGGGAGGGGTTCTCCTTCTGGATCCGC 1347

Qy 841 CGCATAAGGCAGTCGCTCAAGCAGTGCGCATCGATCGAGCCACTGTCGCATCGAGACCACT 900
           || ||||||||||| || |||||| ||||||| || ||| |||| || |||| ||||||
Db 1348 CGTCTGGAGAAGGCCTACTGGGGTGATCACCCCCGAGTGGAGCCGTTGGCGAGGTGGAGGGCC 1407

Qy 901 GAAGGCTTTGGTCTGGAGGGTCCGTTGCCGGCCCAATGCTGTGTGGAGGGGCGCTGGGTCAAA 960
           ||||| ||| || ||||||| |||| || |||| |||| || ||||| |||||| ||| |||
Db 1408 GAGGGCTTCTGGGCAGATCCCCCGAAGCTGCCCCAAGCTGCGCCCACCGCCGGGGCCCGTGGGTCAAG 1467

Qy 961 GCCCAGGGGAGAAGCTGTTTGCTGCCGTGCGGGATGCGCCCATCATT 1020
           || ||||| ||||| ||| ||||| |||| ||||| ||| || ||
Db 1468 GCCCCGGGGGGAAGCTCTTCGGCGAGGTCTTCGGCGGCCCCGTCCTC 1527

Qy 1021 GCCGAAGACCTGGGGGTGATCACCCCCGAGTGAAGCGTTTGCCGATGGGTTC 1080
           || |||||| || ||||| |||||||||||||| |||| |||| ||| |||
Db 1528 GCCGAGGACCTGGGGGTCATCACCCCCGAGTGGAGGCCCGCCGGCCCTT 1587

Qy 1081 CCCGGCCATGAAGATTTTGCAGTTTGCTTTT---TCCGGTGAGGACAACCGCCTTTTTTGCCC 1137
           ||| || |||||||| || |||||| || || ||||| |||||| ||||| ||| |||||||

Db 1588 CCCGGGATGAAGGTCCTGCAGTTCGGCCAGTTGACGACGGGATGGAAAACCCCTTCCTCCCC 1647

Qy 1138 CACAACTACCCCGCCGCACGGCAATGTGGTGGTGTACGCACCACACGACACC 1197
            ||||||||||| || || ||||||||||||| || ||||||||||||||||
Db 1648 CACAACTACCCCTGCCCACGCGGGTGTGGTGCTCTACACCGCACCACGACGACACC 1707

Qy 1198 ACCCTGGGATGGTTCCGCACCGCGCCGGAGGCCGAGGCCGGCCTTCATGCGGCGGGCCTACCTG 1257
           |||| |||| || || || |||| ||||| || ||||| ||| || || ||||| ||||||||
Db 1708 ACCCTGGCTGTACCGGCGGCACGGCCACCCCCAGAGAGGCCTTTCATGCGGCGGTACCTG 1767

Qy 1258 GCCCGCGCTATGGCCATCCGTGTTGTTGTTCGGAATACGAGCTGCGGCGGCTTTGATCGAGCTG 1317
           || || || ||||||||||||| ||||| ||||| ||||| || || ||| |||||||||||
Db 1768 GCGGACTGGGGGGATCACCTTCCCCGGGAAGAGAGAGGAGTGCCCTGGGCCCTGATGCACCTG 1827

Qy 1318 GCCTTCAAAAGCCCGGCCAAGCTGGCTATTGTGCCTTTGCAGGACGTGCTCGGGGCTGGGC 1377
           || |||| ||| || ||||| |||| |||||| ||||| ||||| ||||||||| |||||
Db 1828 GGGATGAAGTCCGTGGCCCGCTCGGCTCCCGTGAGGGCAACGGGGTCTGGGGTGGGCCTGGGC 1887

Qy 1378 CCCGGAGGCCCGCATGAACTTCCCCGCCGACGGGCTGGGGGACAACGGGGTCTGGGCTACGCC 1437
           || ||||| |||| ||| |||||||| || |||| ||||| || |||||||| ||||||||
Db 1888 AGCGAGGCCCGCATGAACTACCCGGGAGGCCCTGGGGAACTGGGCCCTGCCGAGGGCCGCTCTC 1947

Qy 1438 GAAGGCGACCTCGAGCCGCGGTCTGGCCGGCCCGGGCTGCGGCCTGGGGGCTTAGGGCCAGCCAG 1497
           |||| | || |||||||| || ||||||| || ||||| ||||| || |||| |||||||||||
Db 1948 CCGGGGGAGCTTCCCCGGAGCACGGGGCGAGGCGTTTAGGGCCGATGGCCGAGGGCCACGGAA 2007

Qy 1498 CG 1499
           ||
Db 2008 CG 2009

```
CDS      /locus_tag="TTC0847"
         complement(290..1864)
         /locus_tag="TTC0847"
         /EC_number="4.1.3.12"
         /codon_start=1
         /transl_table=11
         /product="2-isopropylmalate synthase"
         /protein_id="AAS81191.1"
         /db_xref="GI:46196776"
         /translation="MVEILDTTLRDGTQGEGVSLSVDDKVAIARRLAAFGIHLIEGW
         PGSNPKDAEFFQRMRGVDLGAARLAAFGATRRKGLAPEEDPSVLALLEAETPVVLEG
         KSWTLHVLEALETSLEENLRMIEETVAFFARRGRRVVYDAEHFFDGYKEDPAYALATL
         EAAFRGCADTLVLCDTNGGSLPEEVAITKAVVERFPGVRVGIHPHNDADLAVANALA
         AVRAGATHVQGTINGYGERCGNLNLTSFLPTLVFKYGIPAISPEKLRELREVSHFVDE
         RANLTPNRRAPYVGEAAFAHKAGVHVSAVLKNPRTYEHIPPEWVGNQRRFLVSDVSGR
         SNLLAKLQELGVDLSKEEARRLLEEVKALEYEGYAFEGAEASFYLLAHRLKGGSLPFS
         VEGFSVFVHGRGLSTAWAEATVRVRVGESLQHTAAESPFGPVSALDRAFRKAVLQFYP
         ELSEVELTDYKVRILSQOEAGTNSGVRVMVEMKRGEERFATVGASENILEASLKALTD
         GYAYALLYPVATPRGA"

gene     complement(1858..2385)
         /locus_tag="TTC0848"

CDS      complement(1858..2385)
         /locus_tag="TTC0848"
         /EC_number="2.3.1.-"
         /codon_start=1
         /transl_table=11
         /product="acetyltransferase"
         /protein_id="AAS81192.1"
         /db_xref="GI:46196777"
         /translation="MVRIRRAGLEDLPGVARVLVDTWRATYRGVVPBAFLEGLSYEGQ
         ABRWAQRLKTPTWPGRLFVAESSGEVVGFATPGDRASGFPGVTAELWAIVLPTWQ
         RKGLGRALFHEGARLLQAEGYGRMLVWVLKENPKGRGFYEHLGGVLLGEREIELGGAK
         LWEVAYGFDLGGHKW"

gene     complement(2411..3973)
         /locus_tag="TTC0849"

CDS      complement(2411..3973)
         /locus_tag="TTC0849"
         /EC_number="4.1.3.12"
         /codon_start=1
         /transl_table=11
         /product="2-isopropylmalate synthase"
         /protein_id="AAS81193.1"
         /db_xref="GI:46196778"
         /translation="MEKRHIRIFDTTLRDGEQSPGVALSLDQKLEIAQALAQLNVDI
         IEAGFPVSGPMFEAVRRIATEVKGPIIAALARTHTLDIDQARAKALEKAEKPRIHVFT
         SASKVHLQYMLRKTEEEVLEMADRMVRYARRYYDEVVRGRDVIISTHTHHDDLGMATANALA
         IEAGATTNIPDTTGYGTPQEYGALIRRIRDEVVRGRDVIISTHTHHDDLGMATANALA
         GVENGAQOIECTINGIGERAGNCALEEVVMALVRRDWYKAYTRINTREIYRVSRLVE
         RYTGMPVPPNKAIVGDNAFAHESGIHQDGVIKHRATYEIMDAELIGRRPAVLVLGKHS
         GRAAPKKALEDLGYKDLSEEEVVKKLFAPKEIAEKKGQALVESERPTSHF
         FOLEHVQFFSGSGLLPATVVKKTPDGERVATHTGDGPVDAVYKAIOEAIGLRPELEL
         YRVBAITGSTEALGQVTVRLRLGELQAVGVGVSPDIIEASALAFLDAAGKLASGRATR
         HPSIEEVHRGV"

gene     complement(4031..5044)
         /locus_tag="TTC0850"

CDS      complement(4031..5044)
         /locus_tag="TTC0850"
         /EC_number="1.1.1.86"
         /codon_start=1
         /transl_table=11
         /product="ketol-acid reductoisomerase"
         /protein_id="AAS81194.1"
         /db_xref="GI:46196779"
         /translation="MKIYYEHDADLGFILGKKVAVLGFGSQGHAHALNLKDSVDVRV
         GLRKGSRSWEKAEAAGLRVLPVABAVREADVVMVLLPDEKQAQVYREEVEPNLKEGGA
         LAFAHGFNVHFGQIKPRKDLDVMWVAPKGPGHLVRSEYQKGSGVPALVAVHQDASGSA
         FPTALAYAKAIGAARAGVIATTFKDETETDLFGEQAVLCGGLTRLIRAGFETLVEAGY
         PPEMAYFEVHEVLVDLIYEAGLKGMYSISNTAEYGDYTRGDLAVPLEETKRRMR
         EILRQIQSGEFAREWMLENQVGSPVLEANRKWAAHPIEEVGSRLRAMMPFLKARVME
         EVG"

gene     complement(5041..5553)
         /locus_tag="TTC0851"

CDS      complement(5041..5553)
         /locus_tag="TTC0851"
         /EC_number="4.1.3.18"
         /codon_start=1
         /transl_table=11
         /product="acetolactate synthase small subunit"
         /protein_id="AAS81195.1"
         /db_xref="GI:46196780"
         /translation="MRHVISVLVQDHPRVLNRITGLFARRGFNLESLAVGTTHQPGLS
         RICLVVSGDDRTLEQVEKQLNRLIEVLKVTDHTEPHVREBLCLVKVHVAGLEERLAVK
         DIQEAFRARVVDVAQKSILELTGDTQKVNSFIEALRPGILEVMRTGAVAMSRGDRV
         LKVREKKEAV"

gene     complement(5550..7238)
         /locus_tag="TTC0852"

CDS      complement(5550..7238)
         /locus_tag="TTC0852"
         /EC_number="4.1.3.18"
         /codon_start=1
         /transl_table=11
         /product="acetolactate synthase large subunit"
         /protein_id="AAS81196.1"
         /db_xref="GI:46196781"
         /translation="MKGAEALLRALEREGVEVIFGHPGGAIMPTYDALYDSPIRHILV
         RHEQGCVHAATAYARASGRVGVVMATSGPGALNLVTGLADAMMDSTPVVAITGNVPRA
         LIGTDAFQEADVTGVTMPITKHNYLVQDVNEIPRVVKEAFHIAATGRPGVLIDIPKD
         VQLAEFTGSFDVKLDLPGYKPTLKGHPKOIERALDALEKAERPVLMVGGGAQHAEL
         LAFAEKTGLPVITTLMGLGAFPGNHPLWLGMPGMHGTVAANRAIHHADVILAIGLRFD
         DRVTGKVSRFAPHAHTIIHVDIDPAEIGKVVRTHIPIVGDSRLVLKELLKGAKALRLA
         AWRELEEWRTRYPLRYRPKPHLQSGVRAFYAATGGHAIVTTGVGQHQMFAAQFPP
         VTRPRSFLTSGGLGTMGVGLPFAIGAKIARPEELVIDFPDGDGSFQMTLQELATVVKY
         LDVKVVILNNGYLGMVRQWQDLFHAKRYSEVYLADSNPDFARLAEAYGIKGVRVERKE
         DLMKGVEAVLNADGPVVAEFKVYHEEGVFPMIPAGGAAEDMILEPEEKEEVGA"

gene     complement(7393..8373)
         /locus_tag="TTC0853"

CDS      complement(7393..8373)
         /locus_tag="TTC0853"
         /EC_number="1.6.4.5"
         /codon_start=1
         /transl_table=11
         /product="thioredoxin reductase"
         /protein_id="AAS81197.1"
         /db_xref="GI:46196782"
         /translation="MVDVIVGAGPVGLCAALWAKRLGLSHLVLERGTVAETVRFPR
         DMVFFSEAKNIEIGGHPLVAQGPKPTRKGALLYQRVAEREGLNVLTYTEVLSIRGRE
         GAFQVLAKDRKGERAFSARYVVVATGYFGNPNRLGVPGEDLPHVLHRYEEAAPFFGRE
         VAVVGGSNSAVEAALDIYRGGAKVTLVHRGKWVRPSVKVWLLPDFENRVKEGSIRAVM
         EARVEAITPEGLWLSRPSGREFLPADFVVLIGVRABDRILREAGVAYEGEGRPWLSPE
         WETSVPGLFAVGSCAYGPDTRSVFIENGREHARAALTAIAARVRGLTPRP"

gene     complement(8531..9001)
         /locus_tag="TTC0854"

CDS      complement(8531..9001)
         /locus_tag="TTC0854"
         /codon_start=1
         /transl_table=11
         /product="prepilin-like protein"
         /protein_id="AAS81198.1"
         /db_xref="GI:46196783"
         /translation="MTRRGLTLLELLLVLGILGVLLGLALPLLSPNRLALDQAARSLA
         TQVTRARLEAIRRNAFVGLQVFTEGAGGYLLFVDQNANRYDPGEFGATHFGQGNWA
         RVRLDPEKSALGNMPLLFDPRGIPAKPITATLVLTSGGATRKVVISQQGRARLE"

gene     complement(9035..9616)
         /locus_tag="TTC0855"

CDS      complement(9035..9616)
         /locus_tag="TTC0855"
         /codon_start=1
         /transl_table=11
         /product="prepilin-like protein"
         /protein_id="AAS81199.1"
         /db_xref="GI:46196784"
         /translation="MRKGLTLVEVLVLVTLVIMGIAFAALLTSQLANLRASAQARFATDA
         KAAAVQVLERRSABVLKSEIVPALLSPYKDAPLDPDNPSGWRSFYFVDYYFSCPTRVA
         PSPKQRGGSVANLRPGLTCSGTETIFGIPVAWDIRGENGILGEGVVTVVTATHPRGP
         KVTLGRRVTCYDVVFSPTQDQPAPCPPPGGGRP"

gene     complement(9613..10314)
```

```
CDS          /locus_tag="TTC0856"
             9613..10314
             /locus_tag="TTC0856"
             /codon_start=1
             /transl_table=11
             /product="prepilin-like protein"
             /protein_id="AAS81200.1"
             /db_xref="GI:46196785"
             /translation="MKRGFTLVEVLVAMAILVVLAVGVRYFASTSELARNTQARSEL
             QDRVRMVMQVVTADLQMAGARYWNSGNQNQAFSLPLPPLJSGSSNMGPKDTLTLYVTSL
             RDLASACRRVDYGFEGDTLRRSDVNATPSSGSDCTTPPNSQPLAEBGMLALDIQYQCS
             DGSRKDTPDGDTDAYPRSAKVTVAGYSLITSVTNPGPASLTTVTGKTLACPQGRACYAL
             TQEYLMPNLKPLPTP"
gene         10325..11989
             /gene="comZ"

Query Match    42.5%;  Score 638.4;  DB 1;  Length 273285;
Best Local Similarity  65.0%;  Pred. No. 5e-98;
Matches  977;  Conservative  0;  Mismatches  516;  Indels  9;  Gaps  2;

Qy       1 ATGCAACTCCAAGCGGCTTTGGAATTTGCTCCACCCACCAGTTTCGGGTCGTGG 54
           ||||| || |  |||| ||| || |  ||| ||  |||||||   ||||||||
Db   55089 ATGGAGCTTCCCCGCGCTTCGGTCTCGCTTCCCACCCACGAGCCTCCCGGCCCCCTAC 55030

Qy      61 GGGATTGGGGCTCTGGGCCGGAGGCCGGAGGGTTTGGACTGGCTGGCCGATGGGGA 120
           |||  |||| ||||||||||||||||  ||||||  |||  ||||||||||
Db   55029 GGGGTCGGGGCGTCCTGGGACGCCCAGGAGGCCGGGGACTTCCTCCGCGTTCCTCGCGG 54970

Qy     121 GCCGCTGGTGGCAGGTCTTACCGCTGGGCGCCTACCAGTTACGGCGACTCGCCGTACCAG 180
           |||||| ||||  |||| ||||||  ||||||||||||||   ||||||||||||||
Db   54969 GGGCGGTACTGGCAGTGCTCCCCTTGGGCCCCACGGGGCTATGGCGACTCCCCCTACCAG 54910

Qy     181 TCCTTCTCGGCTTTTGCGTAACCCGTATTTGGTTGACCCCGAGATGCTGATTGAAAAA 240
           |||||||||| ||||||  |||||  |  ||||||||| ||||||  |||
Db   54909 TCCTTCAGCGGCCTTCGCCGCGGAAAACCCTACCTCATAGACCTGAGGCCCCTCGCGGAAAGG 54850

Qy     241 GGCTGGCTGAAACAAAGCGAAGCGCCCCCGTATCCGACCCCAGCGCGTGGATTATGGC 300
           |||  |  ||  |||||| ||||| ||||||
Db   54849 GGCTACGTG------CGCTGGAGACCCCGGGGCTTCCCCCCAAGGCCCGGGCGTGGACTACGGC 54796

Qy     301 TGGCTTTACCAGACCCGCTGGCCCCGTGTTGCGGCGGGGTTTCGGGCGCAAG 360
           ||||| |||||||||| ||||||||||| |||||||||||||  ||||||||||
Db   54795 CTCCTTACGCCTGGAAGTGGCCCGCCGCCGTGAAGGAGGCCTTCAAGGAAAAG 54736

Qy     361 GCTTCGGCCCAGGATAAGACCCGGGGCCGGGAGGCGGGAGGCCTTCGCCGCCTTCGCCGCCCTTCGCCGCCTGGTGGCTC 420
           |||  ||||||||| ||||||||  ||  |||||||||  || |  |  |||||||| ||
Db   54735 GCCCCCGGGGAGGAGGCGGGAGGCGTTGGCATCCGGACATCGGGTCCCATCTTCGTG 54676

Qy     421 GAAGACTATGCGCTCTTTATGGCCCTCAAGACCCGTTGACGGCAAGCCCTGGAACGAG 480
           |||| ||||| |||  ||  ||||| |||||||||| | |||||
Db   54675 GAGGACTACGCGCTCTTCATGGCCGTTCCCGGAGGGGCGCACGGGGGCCTTCCCTGGAACCGG 54616

Qy     481 TGGAGCCCCGAGCTGCCGCGACCGTGAACCGGCTGCCCAGGCCCGCCCAGGGAGCTG 540
           |||| ||||| | ||||| |||||  ||  | || |||||||  ||| |||||||||
Db   54615 TGGCCCTTCCCCTGCGGGAGCGGGAGGCGGGAGGCCCTTAGGGAGGCGGAAAAGCGCCCTTG 54556

Qy     541 GCCGAGGAGTGGGCCTTTACGAGTGGATTCAGTGGTGCTTTTATCTGGAATGGGCCAG 600
           |||||||| ||| ||||| ||| || |||| || |  |||  |||
Db   54555 GCCGAGGAGGGTGGGCGTCCACGCGTTGCCCCAGTGGCTTCTTCTTCCGCGCCCAGTGGGGGGGC 54496

Qy     601 ACCAAGGCCTATGCGGAGGCCGAGGCGTTGGCATCCAAGGGGATTCAGATTATCGGGATCAGGAC 660
           |||  |||| | |||||||| |||||||  |||| ||| || |||  |||||  |||||||||||
Db   54495 TTGAAGGCCGAGGCCGAGGCGTGGCGGGGGGTGCCCCCGGAGGACATCGGGGACATGGGACATCTTCGTG 54436

Qy     661 GCCTTCGATTCCTCAGATGTCTCGGGCCAACCCGCCAGTACTTCTACCTCGAGGCCGATGGC 720
           |  ||  | | |  |||| |||| ||| || | ||||||||| ||||| |||||  |||
Db   54435 GCCGAGGACTCCGCCCGAGGTCTCGGCCCCACCCCCGAGGTGGTTTCACCCTGGACGAGGAGGGGC 54376

Qy     721 AACCCACCGTGGTTGGCGGGCGTTCCGGCGGGACTACTTCTCCGGAAACCGGCCAGCTCTGG 780
           ||||||||  ||| |||||  |||| ||| ||  ||||| |  |||||||  |||||||||||
Db   54375 CGCCCCACCGTGGTGGCGGGCGTGCCCCCGCCGGACGACTTCTCGGAGACGGGCCAGCGCCTG 54316

Qy     781 GGCAATCCGCTCATCGCTTGGGATGTGATGGAAGGGACAACTTTGCCTGGTCGGTCATTGCC 840
```

```
Db   54315 GGCAACCCCTTACCGCTGGGACGTTTTGGAGCGGGAGGGGGTTCTCCTTCTGGATCCGC 54256

Qy     841 CGGATAAGGCAGTGCGTCAAGCAGTGCCACCTGGTCGCCACACTTCCGCGGGTT 900
           ||   ||| || | ||| |||  |||||  |||||| | ||||| ||  |||||||
Db   54255 CGTCTCGAGAAGCGCCTCGAGCTCTTCACCTGGTGCCGCATAGACCACTTCCGCGGCTTT 54196

Qy     901 GAAGCCTACTGGGAGGTTCCGTTTGGCCGGCCCAATCGTGTGGAGAGGGCGCTGGGTCAAA 960
           |||||| ||||||||||| || ||||||||||| | || | ||  ||||||| ||||||
Db   54195 GAGGCCTACTGGGAGATCCCCACGCGGGTGTCCCCACGGGGTGTGGGGGCGGTGGGTCAAG 54136

Qy     961 GCCCCAGGGGAGAAGCTGTTTCTGCGGGTGCGGGCCCAACTGAGCGATGCGCCCATCATT 1020
           ||||||||||||||| ||| ||  ||||  |||||  ||||||| |  |||  ||| ||
Db   54135 GCCCCGGGGGAGAAGCTCTTCAGAAGATCCAGAAGTCCGAGAGTCTTCGGCGGAGCCGTCCCCGTCCTC 54076

Qy    1021 GCCGAAGACCTGGGGGTGATCACCCCCCAGGTGGAGCGCTTTGCGGATGGCGTTCGGGTC 1080
           |||||||||| |||| || |||||||||||  |||||| |  ||  |||||||  |||||
Db   54075 GCCGAGGACCTGGGGGTCATCACCCCCGAGGTGGAGCGCGACCGCTTCGGCCCT 54016

Qy    1081 CCCGGCGATGAAGATTTTGCAGTTTGCTTT---TCCGGTGAGGACAACGCCTTTTGCC 1137
           |||||||| |||| |  || ||||| ||||   |  ||||||| || |||||  ||||
Db   54015 CCCGGGGATGAAGGTCCCCGCAGTTCGCCTTGACGACGGGATGAAAACCCCTTCCTCCCC 53956

Qy    1138 CACAACTACCCCGCACGGCGCAATGTGGTGTACAGGCGGAACCCACGACAACGACACC 1197
           |||||||||||||| ||||||  || ||||| | |||||||||||||||| ||||||||
Db   53955 CACAACTACCCTGCCCACGGCGGGTGGTGTCACCGCGAGGACTACCCGCACCAACGACACC 53896

Qy    1198 ACCCTGGGATGTTCCGCACCGGCCCTTCATGCGGGCCTACCTG 1257
           ||||| || |||| || |||||||||||||||||| ||| |||
Db   53895 ACCCTGGCTCGTGTACCGGCCACCGCCACGGAGAAGGCCTTCATGGCGGGTACCTG 53836

Qy    1258 GCCCGCTATGGCATCCGTGTGTTTGTGCGGAATACGAGGTCGCGGGCGTTTGATCGAGCTG 1317
           ||||| ||||||||| |  |   |  ||||   ||| |||||||  |  |||| |||||||
Db   53835 GCGGACTGGGGGGATCACCTTCCCCGGAAGAGGAGGAGGTGCCCTGGCCCTGATGCCACCTG 53776

Qy    1318 GCCTTCAAAAGCCCGGCCAAGCTGGCTATTGTGCTATTGTGCCGGCTGCTGCTGCTGGGGC 1377
           || ||||  |||||||||  || |||| ||| ||| ||||| ||||||| ||||| |||||
Db   53775 GGGATGAAGTCCGTGGCCCCGGCCTCTCACCCGTGCAGGACGTGCCTGGCCCTGGGGC 53716

Qy    1378 CCCGAGCCCGCATGAACTTCCCCGGACGGTCGGGGAACTGGGGTGGGCTACGCC 1437
           ||||| |||||||||||| ||||| |||   | |||||||||||  || ||||||
Db   53715 AGCGAGGCCCGCATGAACTACACCCGGAACTACCCGGGAAGGCCTGGGGACTGGGGGCTCTC 53656

Qy    1438 GAAGGCGACTCGAGCCCGGTCTGGGCCGCGGACTGGGGCCTGCCAGCCAGCCAG 1497
           ||||  ||||| | ||| |  |||||| ||   |  |||  |||||||||||||
Db   53655 CCGGGGGAGCTTTCCCCGGAGCACGGGGCGAGGCTTAGGGCGGCATGGGCGAGGCACGGAA 53596

Qy    1498 CG 1499
           ||
Db   53595 CG 53594
```

```
RESULT 6
BD179859
LOCUS       BD179859                1500 bp    DNA     linear   PAT 15-MAY-2003
DEFINITION  Highly thermophilic bacterium-derived protein and gene encoding it.
ACCESSION   BD179859
VERSION     BD179859.1  GI:30790777
KEYWORDS    JP 2002325574-A/350.
SOURCE      Thermus thermophilus
  ORGANISM  Thermus thermophilus
            Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
            Thermus.
REFERENCE   1  (bases 1 to 1500)
  AUTHORS   Kuramitsu,N. and Yokoyama,S.
  TITLE     Highly thermophilic bacterium-derived protein and gene encoding it
  JOURNAL   Patent: JP 2002325574-A 350 12-NOV-2002;
            THE INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH
  COMMENT   OS   Thermus thermophilus
            PN   JP 2002325574-A/350
            PD   12-NOV-2002
            PF   23-FEB-2001 JP 2001116171
            PI   NARUKI KURAMITSU,SHIGEYUKI YOKOYAMA
```

PC   C12N15/09,C12N15/09,C07K14/195,C12N1/15,C12N1/19,C12N1/21, PC
      C12N5/10,
PC   C12N9/88,C12P21/02//(C12N9/88,C12R1:01),(C12N15/09,C12R1:01),
PC   (C12P21/02,C12R1:01),C12N15/00,C12N15/00,(C12N15/00,
PC   C12R1:01)
CC   Highly thermophilic bacterium-derived protein and gene   CC
      encoding it

FH   Key   Location/Qualifiers
FT   CDS   (1)..(1500).
FEATURES   Location/Qualifiers
      source   1..1500
                    /organism="Thermus thermophilus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:274"

ORIGIN

Query Match   42.4%;   Score 636.8;   DB 6;   Length 1500;
Best Local Similarity   65.0%;   Pred. No. 2e-97;
Matches 976;   Conservative 0;   Mismatches 517;   Indels 9;   Gaps 2;

Qy   1 ATGCAACTCCAACGCGCTTTTGGAATTTGCTCCACCCCACCAGTTTTCCGGGTCGCTGG   60
           |||||||||||||||| ||||||||| || || || | | ||||||||||||||
Db   1 ATGGGAGCTTCCCGCGCCTTTGGTCTCTTCCCACCCCAGAGCCTCCCCGCCCCCTAC   60

Qy   61 GGGATTGGGGCTGTCGGGCCGCGAGGCCGAGCGGTTTTGGACTGGCTGGCGATGCGGGA   120
           ||||| ||||||||||| ||||||||||||||||| ||| |||| || ||||||
Db   61 GGCGTCGGCGTCCTGGGCCAGGAGGCCCGGGACTTCCTCCGCTTCCTCAAGGAGGCAGGG   120

Qy   121 GCCCGCTGGTGTGGCAGGTCTTACGCGTGGGCCCTACCAGTTACGGCGACTCGCCGTACCAG   180
           ||| || ||| |||||||| | |||||||| | |||||||| ||||||||||||||||
Db   121 GGGGCGGTACTGCGCAGGTCCTCCTCCCCTTGGGCCCCACGCGGCTATGGCGACTCCCCTACCAG   180

Qy   181 TCCTTCTCGGCTTTGTCCGCGGTAACCCGTATTTGGTTGACCCCGAGATGCTGATTGAAAAA   240
           ||||| || ||||||| ||||| |||||| | |||||| ||||||||||||||| ||||||
Db   181 TCCTTCAGCGGCTTCGGCGCGGAAAACCTCATAGACTTAGAGGCCCCTGAGGGCGGAAAGG   240

Qy   241 GGCTGCTGGAACAAAGCGAAGCGGGCTTCGGGCGGGGTTTCGGGCAAGG   300
           ||||| ||||||||||||||||| ||||| |||||||| ||||
Db   241 --CGCCTGGAGGACCCCCGCTTCCCCCAAGGCCGTGGGGGCTTTCGGGCGTGGGCTACGCGC   294

Qy   301 TGGCTTTACCAGACCCGGCGCTGGCCCCTGTTGCGCGGCGTTTCGGGCAAGG   360
           | || | ||||| |||||||| ||||| || || |||||||||||||||||
Db   295 CTCCTCTACGCGCCAGGATAAGACCCGACTGGAAGCCTTTATCGAGGCCGAGCGCTTCTTGGCTG   354

Qy   361 GCTTCGGCCCAGGATAAGACCCGACTGGAAGCCTTTATCGAGGCGTTCTTGGTG   414
           ||| ||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   355 GCCTCCCCGGAGGAGGCGGGAGCCTTCGCCGCCTTCGCGGAGGCCGTTGGTGGCTC   414

Qy   421 GAAGACTATGGCGTCTCTTTATGGCCCTCAAGACCCGGTTTGACGGACCGGCCGACGAG   480
           ||||| ||||||||||| || ||||| |||||||||||||||||||| ||||| |||
Db   415 GAGGACTACGCCCCTCTTCTCATGGCCCTGAAGGGGGCCGCACGGGGGCCCTTCCCTGGAACCGG   474

Qy   481 TGGAGCCCCGAGCTGCCGCGACCGTGAACCGGCCGTCCGCTGGCCCAGGGCCGGTGAGGAGGCTG   540
           |||| | | || ||| |||||| ||||| ||||| || |||||||||||| |||||
Db   475 TGGCCCCTTCCCCGCCGAAGCGCCCTTACGAGTGCGCCTTTTTATCGGAATGGGGGCCAG   534

Qy   541 GCCGAGGAGGTGGCCCTTACGGGCGGGGGCCGTTCGCGGGACGTGTGGGGGGCC   600
           ||||| |||| |||| | |||||| |||| |||| ||| ||||||||| |||
Db   535 GCCGAGGAGGTGGCCCCCCACCCGGGGGCCACCCGGGACATGCCCCATCTTCGTG   594

Qy   601 ACCAAGGCCTATGCCGGAATCCAAGGGGATTCAGATTATGCCCATCTTGTG   660
           ||| || |||| | ||||| ||||| ||| ||||| |||||||||||||||
Db   595 TTGAAGGCGGAGGCCGAGGCGTTGGCATCCGGATCATCGGGGACATGCCCATCTTCGTG   654

Qy   661 GCCTTGATTCCTCAGATGTCTGGCCCAACCCCAGTACTTCTACCTGAGGCCGATGGC   720
           ||||| || |||||||||| ||||||| ||| |||||||||| || |||| |||||
Db   655 GCCGAGGACTCCGCCGGAGGTCTGGGCCCACCCCAGTGTTTCACCTGGACGAGGAGGGGC   714

Qy   721 AACCCCACGGGTGGTGGGGGCCGTTCGCGGGACTACTTCTCCGAAACCGGCCAGCTCTGG   780
           |||| | |||| ||||||||| | |||| | ||||||||||| ||||||||| ||| || |
Db   715 CGCCCCACGGTGTGGGGTGGGGTGCCCCCGACTACTTCTCCGAGACGACGGCGCGCTGG   774

Qy   781 GGCAATCCGCTATCGCTGGGATGTGATGGAAGAAGGGACAACTTTGCCTGCGTGCATTGCC   840
           ||||||||| |||||| |||| || ||| ||||||||||||| |||| || | |||||| |
Db   775 GGCAACCCCCTTACCGTCTGGGACGTTTGGACGGGGGAGCGGGAGGGGTTCTCCTTCTGGATCCGC   834

Db   841 CGCATAGGACAGTGCGTCAAGCAGTGCCACCTGGTGCCGCATCGGACGAGCCATTCCGCGGTTT   900
           | |||| | |||||| || || ||| | | |||||| ||| ||| || | ||||| |||
Db   835 CGTCTGGAGAGAGGCCCTGGAGCTCTTCCACCTGGTGCGCATAGACCACTTCCCGCGGCTTT   894

Qy   901 GAAGCCTACTGGGAGGTTCGTTGGCCGCCCAATGCTGTGGAGGGGCGCTGGGTCAAA   960
           || ||| | ||| ||| ||| |||||||| ||||| ||| ||||||| || |||| |
Db   895 GAGGGCCTACTGGGAGATCCCGCAAGTGCCCCCACGCGGTGGAGGGGCGCGTGGTCAAG   954

Qy   961 GCCCCAGGGGAGGAAGCTGTTTGCTGCGCGCGTGCGGCCGGCCCAACTGAGCGATGCGGCCCATCATT   1020
           ||| || ||||| ||| | ||| || | ||| || ||||||||| |||| ||| ||| |||| || ||
Db   955 GCCCCGGGGGAGGAAGCTCTTCCAGAAGATCCAGGAGAGTCTTCGGGGAGGTCCCCGTCCTC   1014

Qy   1021 GCCGAAGACCTGGGGGTGATCACCCCCGAGGTGAGCGCTTCGCGGCGTTCGGGGTC   1080
           |||| ||| | || ||||| |||||| | |||| | || || || || ||||||
Db   1015 GCCGAGGACCTGGGGGTCATCACCCCCGAGGTGAGGAGCCCTGCCGACCGCCTTCGGCCTT   1074

Qy   1081 CCCGGCATGAAGATTTTGCAGTTTGCTTTT---TCCGGTGAGGACAACGGCCTTTTTGCCC   1134
           |||| || | |||||||| ||||| ||||| | ||| ||| ||||||||| ||||||
Db   1075 CCGGGGATGAAGGTCCTGCAGTTCGCCTTTGGGCTTCCGGAGGGATGGAAAACCCCTTCCTCCCC   1134

Qy   1138 CACAACTACCCCGCCACGCGGCAATGTGGTGTCTACAGGCACCACC   1197
           ||| ||||||||||| |||||||| || || || |||| ||||||
Db   1135 CACAACTACCCCGCCACCGCGGGGTTGGTGTCTCACCGGCACCACC   1194

Qy   1198 ACCCTGGATGTTCCGCACCGGCCGAGCGGCCTTCATGCGGGCCTACCTG   1257
           ||||||||| |||||||||||||| |||| || ||||| |||||||||||
Db   1195 ACCCTGGGCTGTGTACCGCACCGGCCCACCGGAAGAGGCCTTCATGGCGGGGTACCTG   1254

Qy   1258 GCCCGCTATGGCATCCGTGTTGTTGTCGGGAATACGAGTCGCGGGCGCGTTTGATCGAGCTG   1317
           |||| | || |||| || || | | || ||| |||||| || ||||| || |||| ||||
Db   1255 GCGGACTGGGGGGATCACCTTCCCGGAGGAGAGGAGGAGGTGCCCCTGGGGCCCTGATGCACCTG   1314

Qy   1318 GCTTCAAAGCCCGGCCAAGCTGCTATTGTCTTTGCCTTTGGGGACGTGTTGGGCGTGGGCTACGCC   1377
           ||||| ||| ||||||||||| |||||||| |||| |||| ||||||| |||||| |||| |||
Db   1315 GGGATGAAGTCCGGCCCGATGAGTGACTGACTACCCCGGGGAACTGGGGGGGAACTGGGGCTCCTC   1374

Qy   1378 CCCGAGGCCCAGCTCCCCGGTCTGGCGGGACTCGGCGGGCCTGGCCGAGGCCAGCCAG   1437
           |||||||| |||| | |||| || || |||| || ||||| | |||||||||| ||
Db   1375 AGCGAGGCCGGATGAGAATACACCCGGGAAGGTGCGGCGGCCTGGGCGAGCTGCAGCCAG   1434

Qy   1438 GAAGGCGGACCTCGAGCCCGGTCTGGCGCGGGACTGCGGGAGGCCACGGAA   1497
           || || || ||||| ||| |||| || || ||| ||||||||||||||
Db   1435 CCGGGGGAGCGTTTCCCCGGACGCTTAGGGCCATGGCCGAGGCCACGGAA   1494

Qy   1498 CG   1499

Db   1495 CG   1496

RESULT 7
AX801458
LOCUS   AX801458   1503 bp   DNA   linear   PAT 13-OCT-2003
DEFINITION   Sequence 253 from Patent WO03002728.
ACCESSION   AX801458
VERSION   AX801458.1   GI:37654049
KEYWORDS   .
SOURCE   Thermus thermophilus
  ORGANISM   Thermus thermophilus
             Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
             Thermus.
REFERENCE   1
  AUTHORS   van der Maarel,M.J., Dijkhuizen,L., Binnema,D.J., van der Veen,B.A.
             and Vos,A.
  TITLE   Prevention of retrogradation of starch
  JOURNAL   Patent: WO 03002728-A 253 09-JAN-2003;
             Toegepast-Natuurwetenschappelijk Onderzoek TNO (NL)
FEATURES   Location/Qualifiers
     source   1..1503
                     /organism="Thermus thermophilus"
                     /mol_type="unassigned DNA"

```
CDS             /db_xref="taxon:274"
                1..1503
                ./codon_start=1
                ./transl_table=1
                /product="amylomaltase"
                /protein_id="CAE51852.1"
                /db_xref="GI:37654050"
                /translation="MELPRAFGLLLHPTSLPGPYGVGVLGQEARDFLRPLKEAGGRYW
                QVLPLGPTGYGDSPYQSFSAFAGNPYLIIDLRPLAERGYVRLEDPGFPQGRVDYGLLYA
                WKWPALKEAFRGFEKASPEREAFAAFREREAWWLEDYALFMALKGAHGGLPWNRWP
                LPLRKREEKALRRAKSALAEEVAFHAFTQWLFFRQWGALKAEABALGIRIIGDMPIFV
                AEDSAEVWAHPEWFHLDEEGRPTVVAGVPPDYFSETGQRWGNRLPLYRWDVLREGFSFW
                IRRLEKALELFHLVRIDHFRGFEAYWEIPASCPTAVEGRWVKAPGEKLFQKIQVFGE
                VPVLAEDLGVITPEVEALRDRFGLPGMKVLQFAPDDGMENPLPHNYPAHGRVVVYTG
                THDNDTTLGWYRTATPHEKAFMARYLADWGITFREEEEVPWALMHLGMKSVARLAVYP
                VQDVLALGSEARMNYPGRPSGNWAWRLLPGELSPEHGARLRAMAEATERL"
ORIGIN
```

Query Match 42.4%; Score 636.8; DB 6; Length 1503;
Best Local Similarity 65.0%; Pred. No. 2e-97;
Matches 976; Conservative 0; Mismatches 517; Indels 9; Gaps 2;

```
Qy    1 ATGCAACTCCAACGCGCTTTTGGAATTTTGCTCACCCCACCAGTTTCCGGGTCGCTGG 60
          |||  ||| |||||||| |||  |   | ||  | |  |||||||| ||  |    |||
Db    1 ATGGAGCTTCCCCGGGCCTTTCGGTCTCGCTTCCACCCCACGAGCTTCCCCGGCCCTAC 60

Qy   61 GGGATTGGGGCTCTGGGCGCGAGCGGTGTTTTGGACTGGCTGGCCGATGCGGGA 120
          |||  | || ||||||||  |||   |  |  || ||||  || ||||  |||
Db   61 GGGCTCGCGGCGTCCTGGGCCAGGAGGCCCGGGACTTCCTCCGCTTCCTCAAGGAGGCAGGG 120

Qy  121 GCCCGCTGGTGGCAGGTCTTACCGCTGGGCCCTACGGCGACTCGCGTACCAG 180
          |   | ||||||||||| | ||| ||||||| |||| || ||    ||||||
Db  121 GGGCCGGTACTGGCAGGTCCTGCCCCTTGGGCCCACGGGCTATGGCGACTCCCCCTACCAG 180

Qy  181 TCCTTCTCGGCTTTTGCCGTAACCCGTATTTGGTTGACCCCGAGATGCTGATTGAAAAA 240
          |||||   | || || |||| |   |||||| | ||| || || |||| | |  ||| |
Db  181 TCCTTCAGCGCCTTCGCCGGAAATCCGTATCTCATAGACCTGAGGCCCCTCGCCGAAAGG 240

Qy  241 GGCTGGCTGGAACAAAGGAGCCGGGCGTATCCGCAGCGGCGTGGATTATGGC 300
          ||||||          ||| || || || |||  ||  ||  ||| |  |
Db  241 GGCTACGTG------CGCTGGAGGTGCCGTTCGTGCCCCGGGGTGTCTACACCGGC 294

Qy  301 TGGCTTTACCAGACCCGCTGGCCCGTGTTGCGGGGGTTTCGGGCAAGG 360
          |||| | |||||       | |||  |||  |   |||| | ||  ||
Db  295 CTCCTCTAGCCTGGAAGTGCCGGAGAAGTGGGCCCCTGAAGGAGGCAGGGAGGAAAG 354

Qy  361 GCTTCGGCCAGGATAAGACCGACTGGAAGCCTTTATCGGAACGGCCCGAGCCGTTCTCGGCTG 420
          |||||||      | |   ||||  ||||  |  | |  |  ||    ||| | | |   ||
Db  355 GCCTCTCAGCCTGGACTGGCCCTCGCCCGCCTTCCCGGGAGAGGGGACTGGCCGGCCAGCCAG 414

Qy  421 GAAGACTATGCGCTTCTTTATGGGACCCGGCTGCGCGACCGTGAACCGGTTGACGGGAACAGAG 480
          ||||    |||||  |||  ||||||||||||| |||||  |||| ||  |||||
Db  415 GAGGACTACGCCCCTCTTCATGGGCCCCTGCGACCGTGGGCTCTACCCCGGGAACCGG 474

Qy  481 TGGAGCGCCCAGCGCTGCGCGACCGTGAACCGGTCGCCGAGGAGCTG 540
          ||||||  ||| |||| |||||||||| |||||  |||||||||||
Db  475 TGGCCCTTCTCCGCAGAAGCGCGGGAACCGGCGCCTAGGGAGGCGAAAAGCGCCCTTG 534

Qy  541 GCCCAGGAGCGTGGCCCTTTACGAGTGGCTTTTTATCGGAATGGGGCCCAG 600
          ||| ||||| | ||| || | || ||  | |  ||    | ||||  ||||
Db  535 GCCGAGGAGGTGGCCCTTCCCAGCCGCCTGCGCTCTTTCCGCCAGTGGGGGGCC 594

Qy  601 ACCAAGGCCTATGCGAATCCAAGGGATTCAAGATTATCGGCGATATGCCCATCTTTGTG 660
          |||| ||||  |  ||| || || || |    | | |||| || ||     ||  | ||
Db  595 TTGAAGGCCGAGGCCGTTGGAGGCGTTGGGCATCCGGGATCATCGGGGACATGCCCATCTTCGTG 654

Qy  661 GCCTTCGATTCTCAGATGTCTCGGCCAGCCAACCCGGCTACTTCTACCTGAGGCGCGATGGC 720
          ||| || ||  | || ||  ||||||  |   |||||| ||||| | ||||  | ||
Db  655 GCCGAGGACTCCGCCGAGGTCTGGGCCCACCCCGAGTGGTTCCACCTGGACGAGGAGGGC 714

Qy  721 AACCCCACGGTGGTGGCGGCGTTCGGCGGGGTGCCCCCGGTGTCCGG 780
          || ||||||  ||||||||| |  || ||   || |||  | | || 
Db  715 CGCCCCACGGTGGTTGCGGGTGTCCCCCCAGACTTCTCCGAGACGGGGCAGCGCTGG 774

Qy  781 GGCAATCCGCTCTATCGCTGGGATGTGATGAAAGGACAACTTTGCTGGTGCATTGCC 840
          ||||   |  |||  ||||  | |||| |  | ||   ||||  |  | ||||| ||
Db  775 GGCAACCCCCTTCCCTACCGCTGGGACGTTTTGGAGGGCGGAGGGGTTCTCCTTCTGGATCCGC 834

Qy  841 CGCATAAGGCCAGTCGCTCAAGCAGTGCCACCTGGTCGCATCGCATGGTCGCGCGGGGTT 900
          || |||| ||||||||| |||| ||| |||||  | || |  | ||||  ||  |  | |
Db  835 CGTCTGGAGAAGGCCCTCGAGCTGTTCCACCTGGTGCGGATCGACCACTTCCGCGGGTTT 894

Qy  901 GAAGGCCTACTGGAGGTTCCGTTTGGCCGGCCAATGTGTGGAGGGGCGCTGGTCAAA 960
          ||| || || ||| |  |||| | || ||  ||| | | ||||||  | |||| |||
Db  895 GAGGCCTACTGGGAGATCCCGGCAAGCTGCCCCACGGCGGTGGAGGGCCGCTGGGTCAAG 954

Qy  961 GCCCCAGGGGAGAAGCTGTTTCGTGCCGCGTGCGGGCCCAACTGAGCCGATGCGCCCATCATT 1020
          ||| | || ||||| |||| | | |||| |  ||| |||| ||||| | || |  ||||  |
Db  955 GCCCCGGGGGAGAAGCTCTTCCAGAAGATCCAGGTGTTCGGCGAGGTGCCCGTGCTGGCCGAG 1014

Qy 1021 GCCGAAGACCTGGGGGGTGATCACCCCGAGGTGGGATGGCTTCGGGGTC 1080
          ||| || || |||||   ||||||||||||||| || | ||  | |||
Db 1015 GCCGAGGACCTGGGGGTCATCACCCCCGAGGTGGAGGCCCTTCGCGACCGTTTCGG 1074

Qy 1081 CCCGGGCATGAAGATTTTGCAGTTGCTTTT---TCCGGTGAGGACAACGCCCTTTTGCCC 1137
          |||||| ||||| |   |  |||| || ||   || || || ||  |  ||  |||| ||
Db 1075 CCCGGGATGAAGGTCCTGCAGTTCGCTCCGGACGATGGGATGGAAAACCCCCTTCCCCAC 1134

Qy 1138 CACAACTACCCCGCGCACGGCAACCCACGACAACGACACC 1197
          |||||||||||||| ||||  || |||||| || || |||
Db 1135 CACAACTACCCTGCCCACGGCCGGGTGGTGGTCTACACCGGCACCCACGACAACGACACC 1194

Qy 1198 ACCCTGGGATGGTTCCGGCACCGGCCGAGCCGGCCTTCATGCGGGCCTACCTG 1257
          ||||  |||| | | |  || || |||  ||| | ||||| | ||| ||||||
Db 1195 ACCCTGGGCTGTTACCGGACGGCCACACCCCACGAGAAGGCCTTCATGGCGCGGGTACCTG 1254

Qy 1258 GCCCGCTATGGCATGCCGTGTTGTGTCGGAGAATACGAGGTCGCGGGCGTTTGATGCGAGCTG 1317
          ||  || |  || || |  |   |  | |||| ||| |||| | ||  | |  ||| |||||
Db 1255 GCGGACTGGGGGATCACCTTCCGCGAGGAGGAAGAGGAGGTGCCCTGGGCCCTGATGCACCTG 1314

Qy 1318 GCCTTCAAAAGCCCCGGCCAAGCTGCTGCTATTGTGCCTTTGCAGGACGTGCTGGGGCTGGGC 1377
          ||  | ||| |   ||| ||||| | ||  | | || || || | ||| | || ||  |||||
Db 1315 GGGATGAAGTCCGTGGCCCGCCTCGCCGTCTACCCGGTGCAGGACGTGCTCGCCCTGGGC 1374

Qy 1378 CCGGAGGCCCCATGAACTTCCCCCGGAGGGGACACGGCTGGGGCGCTGGCTACGCC 1437
          | ||||||| | | |  | || ||||||  | | |||| |||| |  | | |||||
Db 1375 AGCGAGGCCCGGATGAACTACCCGGGACGGCCCTCGGGGAACTGGGCCTGGCGGCTG 1434

Qy 1438 GAAGGGACCTGCGAGGGCCGTGCGGGCCCCTGCCGCCAGCCAG 1497
          || ||||| || |||| |||  | |||| ||||  ||| |||
Db 1435 CCGGGGGAGCTTTCCCCGGAGCACGGGGCACGGGCTTAGGGCCATGGCCGAGGCCACGGAA 1494

Qy 1498 CG 1499
          ||
Db 1495 CG 1496
```

RESULT 8
AP008226_11/c
WPCOMMENT

Sequence split into 19 fragments   LOCUS AP008226   Accession AP008226

Fragment Name	Begin	End
AP008226_00	1	110000
AP008226_01	100001	210000
AP008226_02	200001	310000
AP008226_03	300001	410000
AP008226_04	400001	510000
AP008226_05	500001	610000
AP008226_06	600001	710000
AP008226_07	700001	810000
AP008226_08	800001	910000
AP008226_09	900001	1010000
AP008226_10	1000001	1110000
AP008226_11	1100001	1210000
AP008226_12	1200001	1310000
AP008226_13	1300001	1410000

us-10-618-976-1.rge

```
AP008226_14    1400001  1510000
AP008226_15    1500001  1610000
AP008226_16    1600001  1710000
AP008226_17    1700001  1810000
AP008226_18    1800001  1849742
Continuation (?12 of 19) of AP008226 from base 1100001 (AP008226 Thermus thermophilus HB8

Query Match           42.4%;   Score 636.8;  DB 1;  Length 110000;
Best Local Similarity 65.0%;   Pred. No. 1.1e-97;
Matches 976;  Conservative 0;  Mismatches 517;  Indels 9;  Gaps 2;
```

```
Qy       1 ATGCAACTCCAACGCGCTTTGGAATTTGCTCCACCCACCAGTTTTCCGGGGTCGCTGG    60
            |||||||||||||||| ||||||||||||||| ||||||||||
Db  103076 ATGGAGCTTCCCGCGCGCTTTCGGTCTGCTTCTCCACCCACCCGAGCGCCCGGCCCCTAC  103017

Qy      61 GGGATTGGGGCTCTGGCCGCGAGGCCGAGCGGTGTTTGGACTGGCTGGCGATGCGGGGA   120
            |||||||||||||||||||||||||||||
Db  103016 GGCGTCGCGGCTGCCTGGGCAGGTCTTACCGCTGGGCCCTACCAGTTACGGGCACTCGCCGTACCAGG  102957

Qy     121 GCCCGCTGGTGGCAGGTCTTACCGCTGGGCCCTACCAGTTACGGGCACTCGCCGTACCAG   180
            ||||||||||||||||||||
Db  102956 GGGGCCGTACCGGCAGTGCCTCCTCCCTGGGCCCCCACGGACCTATGGCGACTCCCCTACCAG  102897

Qy     181 TCCTTCTCGGCTTTTGCCGGTAACCCGTATTTGGTTGACCCCGAGAATGCTGATTGAAAAA   240
            ||||||||||||||||||||||
Db  102896 TCCTTCAGCGGCCTTTCGCCGGAAACCCCTACCTGAGGCCCCTGCGGAAAGG  102837

Qy     241 GGCTGGCTGGAACAAAAGCGAAGCGCCCCCCGCGTATCCGCGTGGATTATGGC   300
            |||||||
Db  102836 GGCTACGTG-----CGCTGGAGGACCCGTGTTGGCGGCGGTTTCGGCGGAGGTTTCGGGCAAGG  102783

Qy     301 TGGCTTTACCAGACCCGCTGCCCCGTTGGCGGCCTTTCGCGGCGGGTTTCGGGCAAGG   360
            ||  || |||||
Db  102782 CTCCTCTACCGCTGGAAGTGCCCCGGCCCTTCGAAGGACTGGAGCGCTTTCAAGGAAAAAG  102723

Qy     361 GCTTCGGCCCAGGATAAGACCCGACTGGGAAGCCTTTATCGAGGCCGAGCGCTTTGGCTG   420
            ||||||| |||||||||||
Db  102722 GCTCCCCGGAGGAGGCGGGACCTTTCGCCGCTTCGCCCTGAAGGGCGAGGAGGAGGCCTTC  102663

Qy     421 GAAGACTATGCGCTCTTTATTGCGCCCTCAAGACCCGTTTGACGCGGCAAGCCTGGAACGAG   480
            ||||||||||||
Db  102662 GAGGACTACGCCCTCTTCATGCCCGTCTTCGAAGGGGCCGCACGGGGGGCTTCCCTGAACCGG  102603

Qy     481 TGGAGCCCGAGCTGCGACCGTGAACCGCTGCCCCGCCTGCGGCCTTAGGGAGGAGCTG   540
            ||||||||||
Db  102602 TGGCCCCTTCCCCTGCCGGAGGCGGAGGAGGAGGAAGAGGCCTTTAGGGAGGAGCGAAAAGCGCCTTG  102543

Qy     541 GCCGAGGAGGTGGCCCTTTAGGAGTGGATTCAGTGGGTCTTTTTATCTGGAATGGGCCAG   600
            ||||||||||||||||||||
Db  102542 GCCGAGGAGGTGGCCTTCCACGCCTTCACCCAGTGGCTCTTCTCCGCCAGTGGGGGGGCC  102483

Qy     601 ACCAAGGCCTATGCCGAATCCAAGGGGATTCAGATTATCGGCGATATGCCCATCTTTGTG   660
            ||  ||||||| ||||||||
Db  102482 TTGAAGGCCGAGGCCGGAGCCGTTCGGGCATCCGGATCATCGGGGACACTTCCGTTTCGTG  102423

Qy     661 GCCTTCGATTCCTCAGATGTCTGGGGCCAACCCAGTACTTCTACCTCGAGGCCGATGGC   720
            ||||||||
Db  102422 GCCGAGACTCCGCCGGAGGCGTCGGCTCCGGGGTTTCACCTGGAGCGAGGACGAGGAGGGC  102363

Qy     721 AACCCCACGGTGGTGCGGGCCTTCGGGCGGCCCCCCGGACTACTTCTCGGAGACGCGCTGG   780
            ||||| ||||||||
Db  102362 CGCCCCCACGGTGGTGGGGCCCCAGACTACTTCTCGGAGACGCGGGGCCAGCCGCGCTGG  102303

Qy     781 GGCAATCGCTCTATCGTCGGGATGATGTGATGGAAAGGACAACTTTGCCTGTCGCATTGCC   840
            ||||||
Db  102302 GGCAACCCCTTTACCGTCGGACGTTTTGGAGCGGGAGGGGTTCTCTTCTGGATCCGC  102243

Qy     841 CGCATAGGCAGTCGCTCAAGCAGTGCCAGCGTCGTCCATCGGCCACCTTCCGCACGGGTTT   900
            ||||||||||||||||||
Db  102242 CGTCTGAGAAGGCCCTGGAGCTCTTCGAGCTTCCACCTGTGCCCATAGACCACTTCCGCCGGCTTT  102183

Qy     901 GAAGCCTACTGGGAGGTTCCGTTTGGCCCGGCCCCAATGCTGTGGAGGGGCGCTGGGTCAAA   960
            ||||||||||||||||||
```

```
Db  102182 GAGGCCTACTGGGAGATCTGGGAGAATCCCCACGGCGGTGGGGGCGTGGGGGCGCTGGGTCAAG  102123
Qy     961 GCCCAGGGGAGAAGCTGTTTGCTGCGGGGGCGGGCCCCGATGAGCGGATGCGCCCCCATCATT  1020
            ||||||||||||||||
Db  102122 GCCCCGGGGGGAGAGGGAGAGCTCTTCCAGAAGATCCAGAGTGGAGGCTTTCGGCGGATGCGCCGGGTCCTC  102063

Qy    1021 GCCGAAGACCTGGGGTGATCACCCCGAGTGGACCCTTGCGCGATGGCTTCGGGGTC  1080
            |||||||||||||||||||
Db  102062 GCCGAGGACCTGGGGGTCATCACCCCGAGGTGGAGGGCCCCGACCGCTTCGGGCTT  102003

Qy    1081 CCCGGCCATGAAGATTTGCAGTTTGCTTTT---TCCGGTGAGGACGACAACGCCTTTTTGCCC  1137
            ||||||
Db  102002 CCCGGGGATGAAGGTCCTGCAGTTCGGCCTTTGACGACGGCGGATGGGAAAACCCCCTTCCTCCCC  101943

Qy    1138 CACAACTACCCCGCCACGCGGCAATGGTGGTGGTTCACAGCCGGAGCGCAACGACACC  1197
            ||||||||||||||
Db  101942 CACAACTACCCTGCCCACGGCCGGGTGGTGGTTCTACACCGGACGCAACGACACC  101883

Qy    1198 ACCCTGGGGATGGTTCCGCACCGGGCCCGGAGGCCGAGCGGCCTTCATGCGGGGCCTACCTG  1257
            ||||||||||||||||
Db  101882 ACCCTCGGCGTGGTACCCGCCACGGGCCACCCCACGGAGAAGAGGCCTTCATGCGGGGCGTACCTG  101823

Qy    1258 GCCCGCTATGGCATCCGTCGTGTTTGTCGGAATACGAGGTCGGCGGCGTTTGATCGAGCTG  1317
            ||||||||||
Db  101822 GCGGACTGGGGGGATCACCTTCCGGGAAGAGGAGGAGTGCCCTGGGCCCTGATGCGGCTG  101763

Qy    1318 GCCTTCAAAAGCCCGGCCAAGCTGGCTATTGTGCCTTTGCCTTGCTGGGGGCTGGGCGGC  1377
            ||||||
Db  101762 GGGATGAAGTCCGTGGCCCGGGTCCTGACCGGTGCAGGACGGGGGGACAACTGGGCGTGGGGCCTACGCC  101703

Qy    1378 CCCGAGCCCGATGAACTTCCCCGGACGCGGGGGGGACAACTACACCCGGGGGACAAGG  1437
            ||||||||
Db  101702 AGCGAGGCCCCGATGAACTACACCCGGGAGAGGCCCCGGCGGAGGCTCGGGGGGAGCAGCCAG  101643

Qy    1438 GAAGGCGACCTCGAGCCGGTCTGGCCGGCGGACTCGGGCCCTGCGGGCCACGGAA  1497
            |||||||||||
Db  101642 CCGGGGGAGCTTTCCCCGGAGCAGCGGGGCGGAGGCGCTTAGGGCGCATGGCCGCCACGAA  101583

Qy    1498 CG  1499
            ||
Db  101582 CG  101581
```

```
RESULT 9
AP008226_12/c
WPCOMMENT
Sequence split into 19 fragments    LOCUS AP008226  Accession AP008226
    Fragment Name      Begin          End
    AP008226_00            1          110000
    AP008226_01       100001          210000
    AP008226_02       200001          310000
    AP008226_03       300001          410000
    AP008226_04       400001          510000
    AP008226_05       500001          610000
    AP008226_06       600001          710000
    AP008226_07       700001          810000
    AP008226_08       800001          910000
    AP008226_09       900001         1010000
    AP008226_10      1000001         1110000
    AP008226_11      1100001         1210000
    AP008226_12      1200001         1310000
    AP008226_13      1300001         1410000
    AP008226_14      1400001         1510000
    AP008226_15      1500001         1610000
    AP008226_16      1600001         1710000
    AP008226_17      1700001         1810000
    AP008226_18      1800001         1849742
Continuation (?13 of 19) of AP008226 from base 1200001 (AP008226 Thermus thermophilus H

Query Match           42.4%;   Score 636.8;  DB 1;  Length 110000;
Best Local Similarity 65.0%;   Pred. No. 1.1e-97;
Matches 976;  Conservative 0;  Mismatches 517;  Indels 9;  Gaps 2;
```

```
Qy      1 ATGCAACTCCAAGCGGCTTTCGGAATTTGCTCCACCCACCAGTTTCCGGGTCGCTGG 60
          |||| || |||||||| ||| |:|||||| :||||||||| ||||||||||||
Db     3076 ATGGGAGCTTCCCGCGGCTTTCGGTCTGCTTCGTTCTTCCCACCCACGAGCGCTCCCCGGCCCTAC 3017

Qy     61 GGGATTGGGGCTCTGGCCGCGAGGCCGAGCCGGTTTTGGACTGGCTGGCCGATGCGGGGA 120
          ||| ||||| ||||||||||||||| |||||||||||||||:|| |||||||||||||
Db     3016 GGCGTCGGGGCGTCCTGGGCCCAGGAGCCCGGGACTTCCTCCGGGGACTTCCTCAAGGAGGACAGGG 2957

Qy    121 GCCCGCGTGGTGGCAGGTCTTACCGCGTGGGCCTACCAGTTACGGCGACTCGCCGTACCAG 180
          ||||||| |||||||||||||| |||| |:|||||||||||||||||||||||||||
Db     2956 GGGCGGTACTGGCAGGTCCTCCCCACGGCCGCTATGCGGACGACTCCCCCCTACCAG 2897

Qy    181 TCCTTCTCGGCTTTGCGGTAACCCTATTGGTTGACCCGAGATGCTGATTGAAAAA 240
          ||||| | | ||||||||||||| |||| |||| ||||||||| ||||
Db     2896 TCCTTCAGCGGCTTCGCCGGAAAACCCTGAGCCTGAGGCCCTTCGCGGGAAAGG 2837

Qy    241 GGCTGCTGGAACAAAGGAGAGCGCCCCGCTATCCGACCCAGCGCGTGGATTATGGC 300
          |||||||| ||||||| |||||||||||||||||||||||||||||:|||
Db     2836 GGCTACGTG------CGCTGGAGGACCCCCGTCTTCCCCAAGGCCGGGTTTCGGCAAGG 2783

Qy    301 TGGCTTTACCAGACCCGGCTGCCCCGTGTTCGGCGGCGTTTCGGCGGCTTCGGCAAGG 360
          |||||||||||||||||||||||||||||| |||||||||:||||||:|
Db     2782 CTCCTCTACGGCCTGAAGTGGCCCCGCCTGAAGGAGGCCTTCCGGGGGCTTCAAGGAAAAG 2723

Qy    361 GCTTCGGCCCAGGATAAGACCCGACTGGAAGCCTTTATCGAGGCCGAGCGCTTCTGGCTG 420
          |||||||||||||||| ||||||| |||||||||||||||||||||||||||||
Db     2722 GCTCCCCGGAGGAGGAGGAGGCCTTCGCGCCTCCCCTTCGGGGGAGGAGGAGGCTGGCTC 2663

Qy    421 GAAGACTATGCGGCTCTTTATGGCCCCTCAAGACCCGGTTTGACGGCAAGCGAG 480
          || || ||| |||:|||||||||||||| |||||||:||||||||||
Db     2662 GAGGACTACGCCCTCTCCATGGCCCCTGAAGGGGCGCACGGGGGCTTCCCTGGACCGG 2603

Qy    481 TGGAGCCCCCGAGCGTGCGGACCGTGAACCGGTCGCCGCCTGGAGGAGCTG 540
          ||||||:||| |:||||||||||||||| ||||||||||||||
Db     2602 TGGCCCCTTCCCCTGCCGGAAGCGGGAAGAGAGAAAGCCGGAAAAGCGCCTTG 2543

Qy    541 GCCGAGGAGGTGGCCCTTTACGAGTGGATTCAGTCGCTTTTTATCTGGAATGGGGCCGCAG 600
          |||||||||||||||||| ||||||||| || | |||||||||||
Db     2542 GCCGAGGAGGTGGGCTTCGGGTCTCTTCTTCACCAGTGGCTCTTCTTCCGCGACTGGTGGGGGC 2483

Qy    601 ACCAAGGCCTATGCCGAATCCAAGGGATTCAGATTATCGGCGATATGCCCATCTTTGTG 660
          ||||| |||||||| ||| |||||||||| || ||| ||| |||||||||||||
Db     2482 TTGAAGGCGGAGGCCCAGGCGTTGGGCATCCGACATGCCGGCATCGGGACATGCCCATCTTCGTG 2423

Qy    661 GCCTTCGATTCCTCAGATGTCTGGGCCAACCGCAGTACTTCTACCTCGAGGCCCGATGGC 720
          |||:||:|||||| |||||||||| |||||||||| |||||||||||||:|
Db     2422 GCCGAGGACTCCGCCGGAGTCCGGGCTCCGGGTTTCACCGTGGCGACGAGGAGGGGC 2363

Qy    721 AACCCCACGGTGGTGCGGCGTTCCGCGGCCCCCCGGGACTACTTCTCCGGAAACCGGCAGTCTCTGG 780
          ||| ||| || | ||||||||||||||||||||||| |||||| |||:|
Db     2362 CGCCCCCACCGTGGTGGTGGGGGTGCCCCAGCGGCGGCGGCCAGCCGCTGG 2303

Qy    781 GGCAATCCGCTCATCGCTGGGATGTGATGGAAAGGGACAACTTTGCCTGGTGCATTGCC 840
          |||||:|||| ||| |||| ||:||||||| ||||||:|||||||| || |||:
Db     2302 GGCAACCCCCTTTACCGCTGGGACGTTTTGGAGGCCGGAGGGGGTTCTCCTTTCGGAGGATCCGC 2243

Qy    841 CGCATAAGGCAGTCCTCAAGCAGTGCCACACCTGGTGCGCCGGTTT 900
          ||| || |||||||| ||||| ||||| |||:||||||||
Db     2242 CGTCTGGAGAAGGCCCTGAGCTCTTCACCACCCGTGTGCGCGGCTTTT 2183

Qy    901 GAAGCCTACTGGGAGGTTCCGTTTGGCCCAATGTCTGTGGAGGGGCGCTGGGTCAA 960
          ||||||||||||| ||||||||||||||||||||| |||||||||||||||||
Db     2182 GAGGCCTACTGGGAGATCCCCAAGCTGCCCCACGCGGGCGGGGGCGGCTGGGTCAG 2123

Qy    961 GCCCCAGGGGAGAAGCTGTTTGCTGCGGTGCGGCCCAACTGAGCGGCGCCATCATT 1020
          ||||||||||| ||||| || |||||||||| |||| |||| || |||||||||
Db     2122 GCCCCGGGGAGAGAGCTCTTCAGAAGATCAGAGGGTCTTCCGGAGGTCCCCGTCCTC 2063

Qy   1021 GCCGAAGACCTGGGGGTGATCACCCCCGAGGTGAGGCGTTTGCGCGATGCGTTCCGGGTTC 1080
          |||||||:||||||||||||||||||||||||||| |||| ||||||||:||
Db     2062 GCCGAGGACCTGGGGGTCATCACCCCCGAGGTGAGGCGCCCTTCGGCCTT 2003

Qy   1081 CCCGGCATGAAGATTTGCAGTTTGCTTTT---TCCGGTGAGGACAACGGCCTTTTTGCCC 1137
```

```
Qy    1  ATGCAACTCCAACGCGCTTTGGAATTTGCTCCACCCACCAGTTTCCGGGTCGCTGG   60
         ||| || || |||| || ||| |||| ||||| ||||||||||| ||| |||| |||
Db    1  ATGGAGCTTCCGCGGCGCCTATGGCCTGCTTCTGCACCCACCAGCTTCCCGGACCCTAT  60

Qy   61  GGGATTGGGGCTCTGGCCGCGAGGCCGAGCTGGCTGGCCGATGCGGGA            120
         ||| || ||| ||| ||| | ||||| |  |||||| || |||| |||
Db   61  GGGGTTGGGCGTCCGTGGGGAGGAGGCTTTTGCCTTCCTCCGATTCCTCGGGGGAGTCTGGG  120

Qy  121  GCCCGCTGGTGGCAGGTCTTACCGCTGGGGCCTACCGCTCGCCGTACCAG          180
         ||||| || || ||| || ||   || |||||||  |||||| |||||||
Db  121  GCCCGCTACTGGCAGGTTTTGTGCCCTTAGGGCCCACACGGATACGGAGACTCCCCTACCAG  180

Qy  181  TCCTTCTCGGCTTTTCCGGTAACCCGTATTTGGTTGACCCCGAGATGCTGATTGAAAAA    240
         ||| |||  ||||||||| |||     ||   |||| |||||  || | | |||||
Db  181  TCCTTCAGCGCCCTTTGCCGGGGAACCCCTACCTCATCGACC------TCCGGCTAAGGCGG  234

Qy  241  GGCTGGCTGAACAAAGGCGAAGCGCCCCCGCCAGCGCGTGGATTATGGC           300
         ||||| ||||| |  |||  |||| |||| |||||  ||||  || |||
Db  235  AGGCGGGGTACCTCCGCTCGAGACGCCCTCGCGGGCCCGGTTTCGGGGTTTCGCGGGGTTT  294

Qy  301  TGGCTTTACCAGACCCGCTCGCCCTGTTCGCGGGGGCTTTCGCGGGGTTTTCCAGGAGCGGG    360
         ||||| || |||| |||
Db  301  TGGCTCTACCGCTGAAATGCCGGGCTTAGGGAAGGCCTTTCCGGGGGTTTCCAGGAGCGGG  354

Qy  361  GCTTCGGCCCAGGATAAGACCGACTGGAAGCCTTTATCGAGGGCCGAGCCGCTTCTCGGCTG    420
         ||  ||||| |||||||| ||| ||||||||||| ||| ||||||||| ||||| |||||
Db  355  GCTAGCCGGGAGGAGAAAGAGGCCGTTTCAGCTTTTGGGACCCGGGAAAGGAGGAGTTGGTTG  414

Qy  421  GAAGACTATGCGCTCTTTATGGCCCTCAAGACCCGGTTTGACGCGTTGACAAGCCCGTGGAACGAG    480
         | ||| || ||| | || || || ||||| || ||| ||| || ||||| ||||  || ||
Db  415  GATGACTACACCCCTCTTCATGGCCTTCATGGCCCGGGTTGAAGCCTTGACCCGGGAACCGC  474

Qy  481  TGGAGCCCCGAGCTGGCGGACCTGAACCGGCCGCCTGCCCAGGGCCCGTGAGGAGCTG      540
         || ||||| |||  |||| || |||| |||||| | ||| ||||| || || || |
Db  475  TGGCCCATGCCCTACGCCCTGAGCGGAGGAGAAAAGGCTTTAAGGGAAGGCATCTTTGGCCCTC  534

Qy  541  GCCGAGGAGGTGGCCCTTTACGAGTGGATTCAGTGGCTTTTTTATCGGCGATAATGCCCCAG    600
         |||| ||| |||||| |||  || ||| ||| ||| || |||  |  || ||||| |||||
Db  535  TCCCAAGAGGTGGCCTTCCCACGCCTGGACCCCAGTGGTGTTTTTTCTTCCGGCAGTGGCAGGCC  594

Qy  601  ACCAAGGCCTATGCGCCAATCCAAGGGGATTCAGATTATCGGCGATATGCCGATATCTTTGTG    660
         || |||||| | || |||| |||||||   || |||| | || || || |||  ||||
Db  595  TTGAGGGAGGCGGCGGCCGGCCTTGGGCCTCTCGTGATCGGCGCATGCCCATCCCCATCTTCGTG  654

Qy  661  GCCTTTCGATTCCTCAGATGCTGTGGCGGCCCAACCCAGTACTTCTACCTCGAGGCCGATGGC    720
         || || | |||||| |||| ||||| | || ||| || |  || |||| |||| |||||
Db  655  GCCGAGGACTCCGCCGAGGTGGGCGGGCTGGTTCCACCTGGTTCCACCTGGATGAGGAGGGA  714

Qy  721  AACCCCACGGTGGTGGCGTGGGGGGGTTCCCGCGGGACTACTTCTCCGGAAACCGGCCAGCTCTGG    780
         |||||||| ||||| |||  ||||| ||  ||| || || || ||||| || ||| |||||
Db  715  AGGCCCACGGTGGTGGTGGCCGGGGTGCGGGTTCCCGCCCGGATTAATGCACTTTTTCCGGAAAACGG  774

Qy  781  GGCAATCCGCTCTATCGCTGGGATGTGATGAAAAGGACAACTTTGCCTGGTGCATTGCC    840
         || ||||| ||  ||  || ||  || || |||| ||  ||||  || |||| |||||
Db  775  GGGAATCCCCGTACCCGTGGGAGGTGCTGGAGGTGCTGGAAAGGGAGGAGGGGTTCTCCTTCTTCGGATAGAG  834

Qy  841  CGCATAAGGCAGTCGCTCAAGCAGTGCCACCTGCCACCACTTCCGCGGGTTT    900
         || ||||||| |||||| |||||| |||| |||||| |||
Db  835  CGCCTGAGGAGGGCCTTGGGAGCTTTTCCACCTGGTGCGCATCGAACCACCTTCCGGGGGCTTT  894

Qy  901  GAAGCCTAGTCCGGTCCGTTCCGTTGGCGCGGCCCAATCGTGTGGAGGGGGCGCTGGGTCAAA    960
         ||||| || | ||| ||||| ||||||||||| ||||  || |||||||||||  || |||
Db  895  GAAGCCTACTGGGGAGGAGAAAGCCGTGTTTGCTCGGTGCGGCCCAATGCAGCCAGTGGAGGTGAAG  954

Qy  961  GCCCCAGGGGGGAGAAAGTTCTGTTTGCTGCTGCGGTGCGGCGCGATGCGCCCATCATT   1020
         |||||||| |||| ||| ||| || ||  |  || ||||| || || |||| ||||
Db  955  GCCCCGGGGGAGAGTTGTTCTTTCAGAGAGATTCAGGAAACCTTGCGCCGGGTGCCCATCCTG  1014

Qy 1021  GCCGAAGACCTGGGGGTGATCACCCCCGAGGTGGAGGCTTTGCGCGATGGCTTCGGGTTC   1080
         ||||| |||||| |||||| ||||| ||||||||||| |||  | || || | |||
Db 1015  GCGGAGGACCTGGGGGTGATCACCCCCGAGGTGGAGGCCCCTGAGGGACCCGGTTTTGGCCTT  1074

Qy 1081  CCCGGCGATGAAGAGATTTGCAGTTTGCTTTTTCCGGTG---AGGACAACGCCCTTTTTCGCCC   1137
         ||||| |||| || ||||||||||||||  |||  |||      ||| ||||| ||||  |||
Db 1075  CCGGGGGATGAAGGTCCTGCGAGTTCGCAGTTCGCAGTTCGGCCTTTGACGGGGGGGATGGGAAAACCCCTTCCTGCCC  1134

Qy 1138  CACAACTACCCCGCCGGCAATGTGGTGTGTACAGCGGAGATGGCGAAACCCACGACAACGACACC   1197
         |||||||||||||  |||  ||||  || ||
Db 1135  CACAACTACCCCTCCCACCGGCGCGGGTGTGGTTCTACACCGGCGCACCGGACGACAACGACACC  1194

Qy 1198  ACCCTGGGATGGTTCCGCACCGGCCGAGGCCCTTCATGCGGGCCTACCTG   1257
         || || ||||||  |||||||  |||||  |  |||||||||||  ||||
Db 1195  ACCTTGGGCTGGTACCGCACCGGCCTACCCCCCACGAGCGGGGGCTTCCTTGGGGGCTACCTG  1254

Qy 1258  GCCCGCTATGGCATCCGTGTTGTTGTCGGAATACGAGGTCGGCGGCGCTTTGATCGGAGCTG   1317
         ||||| |||||| || ||||  | ||| ||||  |||  |||
Db 1255  GCGGAGTGGGGGATTGGTTCCAGAGAGGGGAGCAGGAGATCCCTGGGCCCTCATGCCACCTG  1314

Qy 1318  GCCTTCAAAAGCCCGGCCAAGCTGGCCTATTGTGCCTTTGCAGGACGTGCTCGGGGCTGGGGC   1377
         || |||||| ||||||| || || ||| ||  ||||| ||||||||
Db 1315  GGGATGAAGTCGGGTGGCCAGGTGGGCCGTCTACCCGTGGGGGAGATGTGTTGGCGTTGGA  1374

Qy 1378  CCCGGAGCCCGCATGAACTTCCCCGGACGCGGTGGGGGACAACTGGGGCTTAGG   1437
         ||||| || | ||| || | ||||||||| ||| |  |||| ||| |||  ||
Db 1375  AGCGAGGCCCGGATGAACTACCCTGGGCGCGGGCCCCTCGGGCCAACTGGGCTTAGG  1434

Qy 1438  GAAGGCACCTCGAGCCCGGTCTCGGCCGCGGGGACTGCGGGGACCTGGGGCCGAGGCCA   1492
         || || ||| || || ||| | ||| ||||||
Db 1435  CCCGGCCAGTCCTCTGCCCGAGCGCACGGGGGAACCGCTTCGCTGGTGATGGGCCGAGGCCA  1489
```

RESULT 11
AE002007

```
LOCUS       AE002007               10747 bp    DNA     linear   BCT 22-NOV-1999
DEFINITION  Deinococcus radiodurans R1 section 144 of 229 of the complete
            chromosome 1.
ACCESSION   AE002007 AE000513
VERSION     AE002007.1  GI:6459402
KEYWORDS    .
SOURCE      Deinococcus radiodurans R1
  ORGANISM  Deinococcus radiodurans R1
            Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
            Deinococcaceae; Deinococcus.
REFERENCE   1  (bases 1 to 10747)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Fraser,C.M. et al.
  TITLE     Genome sequence of the radioresistant bacterium Deinococcus
            radiodurans R1
  JOURNAL   Science 286 (5444), 1571-1577 (1999)
  PUBMED    10567266
REFERENCE   2  (bases 1 to 10747)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,B.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Minton,K.W.,
            Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S.,
            Smith,H.O., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1..10747
                     /organism="Deinococcus radiodurans R1"
                     /mol_type="genomic DNA"
                     /strain="R1"
                     /db_xref="taxon:243230"
                     /chromosome="1"
     gene            125..2242
                     /gene="DR1635"
     CDS             125..2242
                     /gene="DR1635"
```

```
gene
CDS          /note="similar to GB:AL009126 percent identity: 54.60;
             identified by sequence similarity; putative"
             /codon_start=1
             /transl_table=11
             /product="ABC transporter, ATP-binding protein, EF-3
             family"
             /protein_id="AAF11195.1"
             /db_xref="GI:6459403"
             /translation="MTAGRACAWTCTACACAASNGSAGNARPTRSSGPARTRREGERW
             LIRPSLREVSFTGVPCTGDRAGAVRIIGPVLIALSHAAKEYGPLTVLQDITFAVEP
             GDRVGLVGRNGAGKSTLRLLTGQLLPDGGEVRHAPNVRARALQQDPVFPAGATIDTV
             LESAFRDLDALEABLAAAAAAMSGSSERAILHEALLEQFARRGGYERRSRKEQVTLA
             FGFRGREHDPVTGLSGGERTRLGLAALLVENPDVLLLDEPTNHLDIVMEWLEGFLGR
             YPGAVLIVISHDRQFLDNVTKETAYLRGGTLKVYAGGYTKFREQLDLELEQQAALFAQQ
             SKOIADLQSSADRMKIWGLGMSKLARRAKAMQARVDRMQKGAVAAPPPEERTRITFH
             APESGEVVLDARHLTRRMGERTLFADLLDQIRQGBRVAIIGRNGAGKTLLRTLILGLD
             PSDDPRGRVLTGARVSAGYYDQALRGVDPSQTLYDVAREYVQKDHEAHTLLGTFMFPF
             DHHDKPARILSGGERARLALLRLAQEDHNLLVLDEPTNHLDMEMLEALEDALAAYTGT
             LVMVSHDRDRAFIEGLADRIWLIEDGQFYEYPGWEDYRAKHRSAAELEAEAARTVAAATP
             KVALAPKAKGLWHLKREVEAIEADIARLEAELEGAQRALAEADMNADFTALGQAAHEL
             EVQLEDKMALWGBKQGEVEAKGG"

gene         complement(2246..2614)
             /gene="DR1636"
CDS          complement(2246..2614)
             /gene="DR1636"
             /note="similar to GB:AL009126 percent identity: 72.34;
             identified by sequence similarity; putative"
             /codon_start=1
             /transl_table=11
             /product="conserved hypothetical protein"
             /protein_id="AAF11196.1"
             /db_xref="GI:6459404"
             /translation="MRRARQFLSLAAPPVYSRRFMTAPVLEIAMLNVRPGQTAEFEAAF
             AQAGRLIASIPGYRRHELQRCLEDAHKYVLLVWDTLEAHTVGFRGSSAEYGQWRALLH
             HFYDPFPVVEHFVPVTLPNS"

gene         4271..4828
             /gene="DR1638"
CDS          4271..4828
             /gene="DR1638"
             /note="similar to GB:AL009126 percent identity: 64.37;
             identified by sequence similarity; putative"
             /codon_start=1
             /transl_table=11
             /product="conserved hypothetical protein"
             /protein_id="AAF11197.1"
             /db_xref="GI:6459405"
             /translation="MKLYTRTGDQGQTGLYGADRVSKAHPRVEAYGTVDELNSALGLA
             RAHSTDAALDTDLEYLQNALFDVGADLATRSGTPYEKNLVRIDAEDVAFLEAMIDRYQ
             EAAPPFTGFVHPGGTPAAAALQLARAVARRAERDVIRLLDVEEANAQVQVYLNRVSDL
             LFIMARAVNARSGLSEEAWTVKKRR"

gene         4916..5236
             /gene="DR1639"
CDS          4916..5236
             /gene="DR1639"
             /note="identified by Glimmer2; putative"
             /codon_start=1
             /transl_table=11
             /product="hypothetical protein"
             /protein_id="AAF11200.1"
             /db_xref="GI:6459408"
             /translation="MNALVLLAPLPLALLLGSAALPSGVQVVNHSGAGVQNRVCRA
             GECWTKDRLASGERWLVTSEAGQAPVRLTVGGAPATGQVTAQGRGRGVQQLMLERDSRAA
             AGRQ"

gene         5301..5492
             /gene="DR1640"
CDS          5301..5492
             /gene="DR1640"
             /note="identified by Glimmer2; putative"
             /codon_start=1
             /transl_table=11
             /product="hypothetical protein"
             /protein_id="AAF11201.1"
             /db_xref="GI:6459409"
             /translation="MAATAMKGLRFRPGELRRSEVPRTNLLPGAGQPARARREDRWRA
             WLIVGLSVCLLALLGVTLL"

gene         5546..6040
             /gene="DR1641"
CDS          5546..6040
             /gene="DR1641"
             /note="identified by Glimmer2; putative"
             /codon_start=1
             /transl_table=11
             /product="hypothetical protein"
             /protein_id="AAF11202.1"
             /db_xref="GI:6459410"
             /translation="MTDRSDRSQLAFARLLPRLFRGGQAYVGVEATLSDLSSEQAAQH
             AGGLPHSVAELLAHVNWNRWMLDIIEMGQALPYPHAADTWPQVGAGDWGNVKNEFY
             ELLARIDTHAARPDLASPVNHEETIGELLADFALHTAHHFGQIISVRQALGAWPPAGG
             GDTW"

gene         6037..6528
             /gene="DR1642"
CDS          6037..6528
             /gene="DR1642"
             /note="identified by Glimmer2; putative"
             /codon_start=1
             /transl_table=11
             /product="hypothetical protein"
             /protein_id="AAF11203.1"
             /db_xref="GI:6459411"
             /translation="MSEQPAQDFANAAAQLYLGGVANVTWERAISDLSAADAGRRPAG
             LPHSAAQIVAHVQFWQHYLLAVIAGENPAVAHAAEGWPAGELPAGEWEALRDRFLRD
             AQRLRDLTLDASLTASPDRKGRPYAVALTNSAGHSVYHLGQVVTVRQALGLWPPVGGG
             ATW"

gene         6522..6767
             /gene="DR1643"
CDS          6522..6767
             /gene="DR1643"
             /note="identified by Glimmer2; putative"
             /codon_start=1
             /transl_table=11
             /product="hypothetical protein"
             /protein_id="AAF11204.1"
             /db_xref="GI:6459412"
             /translation="MVSVTELNAGNSSRASHVNRVDDGQRSEIYRRSWWTTPEVSAFM
             LGLGQLFGVDPRDLHATESAYGFWRELGVWTVPQTTA"

gene         6740..7372
             /gene="DR1644"
CDS          6740..7372
             /gene="DR1644"
             /note="identified by Glimmer2; putative"
             /codon_start=1
             /transl_table=11
             /product="hypothetical protein"
             /protein_id="AAF11205.1"
             /db_xref="GI:6459413"
             /translation="MDRAADYGLTEFRDAPALRVEYVSGDPNGRTFAEADTRLLGRQI
             AQVHQQAASFFGDVSGQRRQPLEQFYVRALETVRATAPRYAPQNWAGHWDTVERVFAA
             VPPPRQAAPMLLDWNESQFVWRGGQPYALVDVEASATAPPELDLITFWELLLPAGAPAQ
             AFQAGYREVRPWPDLNPHRAACRLILLALESEGTRDAAQWLAQPAVLETA"

gene         complement(7382..8128)
             /gene="DR1645"
CDS          complement(7382..8128)
             /gene="DR1645"
             /note="similar to GB:M95047 SP:P37457 GB:X52093 PID:153897
             percent identity: 56.50; identified by sequence
             similarity; putative"
             /codon_start=1
             /transl_table=11
             /product="UDP-N-acetyl-D-mannosaminuronic acid
             transferase, putative"
             /protein_id="AAF11198.1"
             /db_xref="GI:6459406"
             /translation="MSLKSLPERLVLFDLPLDPVTLDAALDRLSGWLFESPPTPHTVV
             TLNPBFIVQSRTQPEFVRAMQEADLVTADGVGIVYAARQLYGAEVPRAPGFDLVKGLM
             ERHGPELRVFFLGSKPGVAEQAAQNAVRDYGIQVAGVHHGYCFGPDEDQRVAELVGASG
             AHLLTGMGAGRQBIFNAYWKQIHRAPVAIGCGVDVLAGTAQLAPEWTRKMGVEWV"
```

```
gene        WRVAGDRKRWGRAPRLAQFVAMVQAEKRRK"
            8134...8913
            /gene="DR1646"
CDS         8134...8913
            /gene="DR1646"
            /note="similar to GB:AL123456 percent identity: 49.31;
            identified by sequence similarity; putative"
            /codon_start=1
            /transl_table=11
            /product="nitrogen regulator, putative"
            /protein_id="AAF11199.1"
            /db_xref="GI:6459407"

Query Match     32.8%;   Score 492.8; DB 1;   Length 10747;
Best Local Similarity   60.5%;   Pred. No. 2.6e-73;
Matches  908;   Conservative  0;  Mismatches  567;  Indels  25;  Gaps  5;

Qy     1 ATGCAACTCCAACGCGCTTTTGGAATTTGCTCACCCACCAGTTTCCGGGTCGCTGG  60
           ||||| |||| || |||| ||||||||  ||| ||| ||||||  ||||| |||||||
Db  2680 ATGACACTTCAACGTTCCAGCGGCGTGCTGCCACCCCACCAGCCACCACGACTTAT 2739

Qy    61 GGGATTGGGGCTCTGGGCCGCCGAGCCGAGCGGTTTTGGACTGGCTGGCCGATGCGGGGA 120
           |||||  ||||||| |||||||| |||||| ||  |||||||  ||| || ||  ||||||
Db  2740 GGCATCCGGCGAACTCGGCGCCTCGGCGCGCGCACTTCGTGGACTGGCTGGCTCGGCCGGG 2799

Qy   121 GCCCGCTGGTGCGCAGGTCTTACCGCTGGGCCCTACCAGTTACGGCGACTCGCCGTACCAG 180
           ||| || || || ||| |||||||||| |||||| ||||||||||| |||||| |||||||
Db  2800 CAGGCGCTACTCGGCAGGTCATGCCGCTCGGCCCCACCACCGGCTACGCGGCGACAGCCCTATCAG 2859

Qy   181 TCCTTCTCGGCTTTTGCCGGTAACCCGTATTTGGTTGACCCCGAGATGCTGATTGAAAAA 240
           ||| || | ||||||| || ||   ||| || |||||||||||||||| || ||||| ||
Db  2860 GCGTTTTCGGCGTTCGCGCTGGCGCAACCCTTACCTCATCGACCTGACGCGCGTGAGGGAGAA 2919

Qy   241 GGCTGCGTGGAACAAA------GCGAAGCGCCCCCGCCTGCGTATCGCCAGCGCGTGGAT 294
           ||||||||   |||||      || || |||||| |||| ||||||||| |||||||| ||
Db  2920 GGGCTGCTGCACGACACCGATTTCGAGAGCCGATTCGCCGGCCTTCAACCCGCCGCCGGGTGGAT 2979

Qy   295 TATGGCTGGCTTTACCAGACCCGCTGCCCCGTGCGGGCCTTCGCGGGGTTTCGG 354
           |||||| ||  || | |||| ||| |  | ||||||| |||||||  ||||||
Db  2980 TTCGGCACGCAGTTCGTGTGGCGCAACCAGATGCTCGAGTGCTGACCGACTTTATT 3039

Qy   355 GCAAGGGCTTCGGCCCAGGATAAGACCCGACTGGAGAGCCTTTATCGAGGCGAGCGCTTC 414
           |||||||| ||| ||||  ||| |||  ||||| || |||||||  |||||| |||||||
Db  3040 TACGGCGAACACGCGAACTGACGCGCCGCGCCGTTCGGGGGCGTTCGGGGGCGTTCAAGGGCGCGG 3099

Qy   415 TGGCTCGGAAGACTATCGCGCTCTTATTGCGCTTCAAGACCCCGTTTGACGGCAAGCCCTGG 474
           ||||||||||||||| || ||||||||| ||||||| |||||||||||| || |||||||
Db  3100 TGGCTCGACGACTACGGCCCCTGTTCATGCGCGTGCGCGTCAAGACGGCGCGCCACGGCGCTGCCCTGG 3159

Qy   475 AACGAGTGGAGCCCCGAGCTCGCGCGGACCGTGAACCGGCTGCCCTGGCCCCCGAG 534
           ||||||||||||| ||||||||||| ||||| ||| || || || ||||||||||
Db  3160 AACGCCTGGAGCAGGAGCGCGCTGGCACGCGCTGGGGCGCTGCTCCGGCCCGGCGGAM 3219

Qy   535 GAGCTGGCCGAGGAGGTGGCCCATCAACCGCGTTTACGAGTTCAGTGGATTCAGTGGCTTTTTATCTCGGAAATGG 594
           |||||||| ||||  |||||||||  |||||||||||||||||||||||||||||| ||||||||||||||||||
Db  3220 G-GTTGGAAGGGCCCATCAACCGCGTCAGTTCGTGCAGTTTCTCTTTTCCGCCAGTGG 3278

Qy   595 GGCCAGACCAAGGCCTATGCCGCCATC 654
           ||||||||||||||||| || ||||||
Db  3279 ACAGCGCTGCCTGAGTGACGCCCGTGAGCGCGGCGCGGACCCGCGCGTTGGGCGCCTGATTATCGGGGACATTCCCATT 3338

Qy   655 TTTGTGCCGGTTCCTCAGATGTCTCGGGCCAACC 714
           ||| || ||| |||| ||||||  ||||||||||
Db  3339 TTCGTGGCTCGACTCGACTCCAGCGACGCGCGCCTGGGCCAACCGGGGACGAGTTCTACTTCGACGAC 3398

Qy   715 GATGGCAACCCACCGGTGGTGGCCATCAACCGCGTTCGGCGGGG 774
           ||||| ||||| |||| || |||| ||| | ||  |||||||||
Db  3399 CAGGGCCAGGCCAGCGGGTGGTCGCGGGGCGGCGACTATTTCTCGGAGACCCGGGCAA 3458

Qy   775 CTCTCGGGCCAATCGGTCTCTATCGTCTGCTGGGATGCTGATGGAAGGGACAACTTTGCCTCGGTGC 834
           ||| || |||||||| |||||  || |||||||||||||| ||||||||||||||||||||||||||
Db  3459 CTGTGGGGCCAATCGGTCTCCCTCTACAACTGGCCCAAGATGAAAGAGAAGAACGCGGCTTCGGCGTGTGG 3518
```

```
Qy   835 ATTGCCGCCATAAGGCAGTCGCTCAAGCAGTGCGCACTCGGTGCGCATCGACCACTTCCGC  894
           |||||  | |||  |||||| |||||||| |  |||||||| ||  |||||||||||||||
Db  3519 ATTCGCGCGCTTCCAGGGACGCAGCCTGAAGCTGTACGACACATCCGGGTGGACCCATTTCCGG 3578

Qy   895 GGGTTTGAAGCTACTGGGAGGTTCCGTTTGCCGTTGGCCGGCCCAATGCTGTGGAGGGGCGCTGG  954
           || |||| |||| ||||| |  ||| ||||| || ||||||| |||| | || ||||||||||||
Db  3579 GGCTTTGCCGGGTATTGGGAGAATTCCTTACCCGCGGCGAAACCGCCCATGAAACGGCGGCGGCTGG 3638

Qy   955 GTCAAAGCCCCAGGGAGGAGAAGCTGTTTGCTGCTGCGGCGCCAACTGAGCGATGCCCC 1014
           |||||||||||||| ||  | || |||| |||| |||| |||||||| |||||||||||
Db  3639 GTGCCGGCGCTGGAGCCACGAGATGTTCGAGGCGTGTGCCGGCGCGGTGCGGGCGCGCTGCCC 3698

Qy  1015 ATCATTGCCGAAGACCTGGGGTGATCACCCCGAGGTGGAGGCTTTGCGCGATGGCTTC 1074
           |||||||||||| |||| || || || ||||||||||||| || |  || ||||||||
Db  3699 ATCATCGCCGAGACCTGCGGGTGTGATTACCCCGACGTGCGGGACGTGCGCGACGACTTC 3758

Qy  1075 GGGTTCCCCGCATGAAGATTTGCAGTTTGCTTTTCCGGTGAGGACAACGCCTTTTTG 1134
           ||| || || ||||||| |||  | ||| |||||| |||||| || || |||||||
Db  3759 GGCTTTCCCGGATGCGGCGGTGCTGCGCAATTCGCGGTGCTCGCGGGGCGGGCGGACTTCGCGGTGAAC 3818

Qy  1135 CCC---CACAACTACCCGCCGCACGCAATGTCGTGCGTGAACCCACGACAAC 1191
           |||    |  ||   |||| ||  ||||  || || |||  ||||| ||||||
Db  3819 GACTTTCTGCCCGCCAACATCAAGAAAACCGGGTTGGTCTCACACCGGCACCACCGGCAAC 3878

Qy  1192 GACACCACCCTGGATGGTTCGCACCGCCGAGCCGAGCCGAGCCTTCATGCGGGCC 1251
           |||||| |||  || ||||| ||||||||||||||  |||||| ||| | ||||||
Db  3879 GACACCACCGCGGCGGCGTGGTGGGCCACGAGAGCCCACCAACTTCCGGTC 3938

Qy  1252 TACCTGGCCGCTATGGCATCCGTTGTTGTGCGGAATACGAGGTCGCGGGCGCTTTGATC 1311
           |||||||| || ||||| |      |  ||           ||| || ||| || |||
Db  3939 TACACCAGCAGCGACCCGGGC-----------GAGGACAACTTCGGCTGGCAACTGACT 3986

Qy  1312 GAGCTGGCCTTCAAAAGCCCGGCCAAGCTCGCTATTGTGCCTTTGCAGGACGACTGCTCGGGG 1371
           |||| ||  || |   ||||| || |||| || ||| | ||  |||||||||||| ||  ||
Db  3987 GAGGTGGCGTTCGGCACGCACCCGGCGTGTCGCCGTCGCGAGGACCTGTTGAAC 4046

Qy  1372 CTGGGCCCCGAGGCCCGCCATGAACTTCCCCGACGGCTGGGG---ACAACTGGCGTGG 1428
           ||||| |||||| |||  ||| || ||  |||||| ||||||    ||||||| ||||
Db  4047 CTGGGCAGCGGACGGACCCCATGAACCTGCCCCCGACGACCACACCGGCGCGGGGACGTGG 4106

Qy  1429 CGCTACGGCCGAAGCCGACCTTCGAGCCCGTCGCCCTCACTCGCCGACCGTCTCGGCGGCGAG 1488
           ||||| ||| | ||||||||  |||| |  ||||||| ||  |||||||||||||||  ||||
Db  4107 CGCTACGACCCCGGCGGCGGCCTGGCTACTCGCCGACCGTCTCGGCGGCGCGAGCGTGACGGAG 4166
```

```
RESULT 12
AE017180_12
WPCOMMENT
Sequence split into 39 fragments   LOCUS AE017180 Accession AE017180
Fragment Name    Begin      End
AE017180_00            1     110000
AE017180_01       100001     210000
AE017180_02       200001     310000
AE017180_03       300001     410000
AE017180_04       400001     510000
AE017180_05       500001     610000
AE017180_06       600001     710000
AE017180_07       700001     810000
AE017180_08       800001     910000
AE017180_09       900001    1010000
AE017180_10      1000001    1110000
AE017180_11      1100001    1210000
AE017180_12      1200001    1310000
AE017180_13      1300001    1410000
AE017180_14      1400001    1510000
AE017180_15      1500001    1610000
AE017180_16      1600001    1710000
AE017180_17      1700001    1810000
AE017180_18      1800001    1910000
AE017180_19      1900001    2010000
AE017180_20      2000001    2110000
AE017180_21      2100001    2210000
AE017180_22      2200001    2310000
```

AE017180_23	2300001	2410000
AE017180_24	2400001	2510000
AE017180_25	2500001	2610000
AE017180_26	2600001	2710000
AE017180_27	2700001	2810000
AE017180_28	2800001	2910000
AE017180_29	2900001	3010000
AE017180_30	3000001	3110000
AE017180_31	3100001	3210000
AE017180_32	3200001	3310000
AE017180_33	3300001	3410000
AE017180_34	3400001	3510000
AE017180_35	3500001	3610000
AE017180_36	3600001	3710000
AE017180_37	3700001	3810000
AE017180_38	3800001	3814139

Continuation (13 of 39) of AE017180 from base 1200001 (AE017180 Geobacter sulfurreducens

```
Query Match       28.1%;  Score 421.8;  DB 1;  Length 110000;
Best Local Similarity 58.0%;  Pred. No. 1.7e-61;
Matches 838;  Conservative 0;  Mismatches 577;  Indels 30;  Gaps 4;

Qy       1 ATGCAACTCCAAGCGCGCTTTTGGAATTTGCTCCACCCCACCAGTTTCCGGGTGCTGG 60
             ||  | |  |   ||| | |||| | ||  |   |||| |   |||||| ||| || |
Db   88728 ATGAGACAGAAACGTCGCAGCCGCATCGTCGTTCCATCGACTTCCCTGCCGGGTCCGGGG 88787

Qy      61 GGGATTGGGGCTCTGGGCGCGGCCGAGCGGTTTTGGACTGCTGGCTGGCGATGCGGGA 120
             |||||  |  | ||||| | | ||  | | |||   |  |  ||   || ||| |||| |
Db   88788 GGGATCGGATCGCTCGGCGGAAAGAGTGCCGCCCATTTCATCGATTTCCTCGAAGAGGCGGGGG 88847

Qy     121 GCCCGCTGGTGGCAGTCTTACCGCTGGGCCTACCAGTTACGGCGACTCGCCGTACCAG 180
             | |||| | | | ||| || |||||||||  | | || |||| | |  |||  || | |
Db   88848 CAGAGCCCTGCGCAGGTTCTTCCGCTGGGCCCGAGCGCCTACGGCCAATTCCCCCTATTCC 88907

Qy     181 TCCTTCTCCGGCTTTTGCCGGTAACCCGTATTGGTTGACCCCGAGAGTGCTGATTGAAAAA 240
             | | || |||||||| || || ||||| | |||  ||| ||| | | | |||  || |
Db   88908 TGCTATTCGGCTTTTTGCTGGCAACCCCTGCTGGCGACCCTCGTGATGAG 88967

Qy     241 GGCTGGCTGGAACAAAGCGAAGCGCCCCGCCCAGCCGGTGGATTATGGC 300
             ||   |  ||| | | ||| | |  || || | | |  ||  | |||  |
Db   88968 GGGGACCTGAGCGCCCGAGGAGGCTGTGGCGGAGGTGTCTGCGGGATTCATGTGGACTTTCCG 89027

Qy     301 TGGCTTACCAGACCCGCGCTGCCCCTGTTGCGGCGGGCTTTCGCGGGGTTTCGGGCAAGG 360
             ||| | ||| | |||  | ||||   | || | |||  |||||| ||    ||||| | |
Db   89028 GCTGTGGATGCGGTTCAAGTCACGCGCGTTCCGCTCCGGACCACGCGCCTCCGGGCCCCAC 89087

Qy     361 GCTTCGGCCCAGGATAAGACCCGACTGGAAGCCTTTATCGAGGCCGAGCGGCTTCTGGCTG 420
             | ||  ||| || ||  || |  | ||| || ||||| |  || | | || ||  || |||
Db   89088 GGTGACACGGATCGTAAACAGGAGTTCTGGCACTTTGCGATACTACCCC---CTGGGCTT 89144

Qy     421 GAAGAGCTATGGCCTCTTTATGGCCCTCAAGACCCCGGTTTGACGGCAAGGGCCTGGAACGAG 480
             ||| | | ||| | || | ||| |  ||||| |  ||| ||||| |||| || ||||| |
Db   89145 CACGACGCTGCGCGTTCATGGGCTCTCAAGGAGCATTTCGGCGCAGGGCTGGAATACC 89204

Qy     481 TGGAGCCCCGAGCTGCGCGACCGGTGAACCGGCTGCCCTGGCCAGGGAGGAGCTG 540
             |||| |||| ||||| | ||  | ||| |||||||||||  || |||| | ||||
Db   89205 TGGGCCCAAGGAGTGCCCGGAGTGTCGGCCTCGAAAAGTATTCGGTTCGGCTG 89264

Qy     541 GCCGAGGAGGTGGCCCTTTACGAGTGGATTCAGTGGGCTTTTATCGGGCCAG 600
             |||| |||  | ||| ||||| |||||||||||||| | || ||| |   | |
Db   89265 GGCGTGGCCATTGGCGAACAAGTAAGACCAAGTGCAGTTTTTCCCGTCAGTGCGGACAT 89324

Qy     601 ACCAAGGCCTATGCCGAATCCAAGGGGATTCAGATTATCGGCGATATGCCCATCTTTGTG 660
             ||| ||||| ||||| ||  |||||||||| | |||||||||||||| ||||||||||||
Db   89325 GTCAGGGAATATGCCAACAGCAAGGGGATTGCCGGTTGTCCGCGATATCCCCATCTTTGTG 89384

Qy     661 GCCTTCGATTCCTCAGATGTCTGGGCCCAACCCGCAGTACTTCTACCTCGAGGCCGATGGC 720
             || ||||||||  | ||||||| ||| |  |||| |  ||||| | | ||| |||  | || 
Db   89385 GCTTTCGATTCAACCGATGTCTGACGAATCCGGCGACTTCAAGCTCACGACGAACAAGGGC 89444

Qy     721 AACCCCACGTGGTGCGGCGTTCGCGGGACCGCGGGGTCCCGCCCGCCGGACCGGACCGG 780
             || |||||  ||| | |||| | ||| | || ||||||||||||||||||| |||||||
Db   89445 AAGCCCCACCGTGGTAGCCGTGGTAGCCGGTCCCGGGGGTCCCGCCCGCCGGACTGGACCGG 89504

Qy     781 GGCAATCCGCTCTATCGCTGGGATGTGATGGAGAAAGGACAACTTTGCTGGTCGTGGCATTGCC 840
             || || | ||| || ||||||||  | ||| | ||  | |||| |  |||||| ||| ||  |
Db   89505 GGAACCCTCTCATGACTGGGATGCCATGGCCGTGGCGGATTTCGGCTGGTGGTGGATCGAG 89564

Qy     841 CGCATAAGGCAGTCGCTCAAGCAGTGCCATCGCCATGAGTGTGGCGGCCCATTCCGCGGGTTT 900
             || | | || || || | | |  ||| || | |  || | |||  | |||| ||| | ||| |
Db   89565 CGGTTCCGCGGTTCGTTCGCGCTCCATGATGTGGTCAGGGTGGATCATCACTTCCGAGGGGTTC 89624

Qy     901 GAAGCCTACTGGGAGGTTCCGTTGCCGGCCCAATGTGTGGAGGGGCGCTGGGTCAAA 960
             || |   | |  |||||||||  | |||| |||  |||| | ||| |  ||||| |
Db   89625 GAGGGCGTGCGGAGGTTCCGGCCAAAGAAAAGACGCGGGCCAATGGGTCCCC 89684

Qy     961 GCCCCAGGGGGAGAAGCTGTGTTTGCTGCCTGGCGATGGCGGCGCCCATCATT 1020
             || ||  |||| | | || |||||| ||||| |  || |  |||||   ||||||
Db   89685 GCTCCGGGGACGACCGCACTCTTTGATGCCGTTATCGGGCGGCGTGGGGTCACTGCCGATCATT 89744

Qy    1021 GCCGAAGACCTCGGGGGTGATCACCCCAGGTGGACGTGGAGGCTTGGCGGATGGCTTCGGGTTC 1080
             ||||| ||||| || |||| | |||||  | |||| ||||||  |||| ||  ||| | |||||
Db   89745 GCCGAGGACCTGGGAGTCATTACCCCGGCCGTGGAGGAACTGCGCGGACCGGGTACCGTTTC 89804

Qy    1081 CCCGGCGATGAAGATTTTGCAGTTTGCTT---TTCCGGTGAGACACAACGCCTTTTGCCC 1137
             ||||||||||||||  | ||||| || |||   ||||| |  |  ||   ||  | ||||
Db   89805 CCCGGCGATGAAGATCCTCCCAGTTCGCCTTTGATTCCGGCCGGGCAATCCCTACCCTGCCC 89864

Qy    1138 CACAACTACCCCGGCGACGGGCAATGTGGTGACGCGAATGTGGTGTACAGCGGCAGCGACACC 1197
             |||||||  || | |||  |  |||||||   |||  |||||||| ||||| | ||| |||||
Db   89865 CACAACTTCACCGTGACTGC------GTGGGTCTACACGGCACCCACGACAATGATACC 89918

Qy    1198 ACCCTGGGATGGTTCCCGCACCCGAGCGGCGGGCCTTCATGCGGGCCTACCTG 1257
             ||||  |||||||| |  |  | |||| | |||| | | || ||| |||| | |
Db   89919 ACCCAGGGCTGGTTTCAAAGCTTTCTCGGCGGGGACGAGAAGTTCTGGCCTATACC 89978

Qy    1258 GCCCGCTATGGCATCCGCGTTGTTGTTGTGCGGCAATACGAGGTCGCGGGCGCGCTTTGATCGAGCTG 1317
             | ||| || ||  | |                | |||| |||||| |  ||| | |   | |  ||  |
Db   89979 GACACCACGGGTA----------------AGGAGGTCCATTGGGAGCTGGAGCGTGCTGCGGGCTGGGC 90020

Qy    1318 GCCTTCAAAAGCCCCGGCCAAGCTGCTATTGTGCCTTTGCTGGGGCTGGGGC 1377
             ||| |    |||||| || | |||  |||  ||  ||| | | | |||   |
Db   90021 GGCATGGCCTTCGGTGCCGGTCGGCCGATATGGCGATTTTCCCGGTTCCAGGATGTGTCATTGGGGACTTTGAT 90080

Qy    1378 CCGGAGGCCCCCATGAACTTCCCCGGACGGCTTCGGGGGACAACTGGGGCTACCGC 1437
             ||||| ||||| ||||| | |||||| | | |  |   | || ||| | | ||| |
Db   90081 GGGTCAAGCCGGATGAATCTCCCCGGCAAACACCGCGGGGAACTGGTCGTCGTGGCGGTTATAGT 90140

Qy    1438 GAAGG 1442
             || ||
Db   90141 GAGGG 90145
```

RESULT 13
BA000045_35/c
WPCOMMENT
Sequence split into 47 fragments LOCUS BA000045 Accession BA000045

Fragment Name	Begin	End
BA000045_00	1	110000
BA000045_01	100001	210000
BA000045_02	200001	310000
BA000045_03	300001	410000
BA000045_04	400001	510000
BA000045_05	500001	610000
BA000045_06	600001	710000
BA000045_07	700001	810000
BA000045_08	800001	910000
BA000045_09	900001	1010000
BA000045_10	1000001	1110000
BA000045_11	1100001	1210000
BA000045_12	1200001	1310000
BA000045_13	1300001	1410000
BA000045_14	1400001	1510000
BA000045_15	1500001	1610000
BA000045_16	1600001	1710000
BA000045_17	1700001	1810000

```
BA000045_18   1800001   1910000
BA000045_19   1900001   2010000
BA000045_20   2000001   2110000
BA000045_21   2100001   2210000
BA000045_22   2200001   2310000
BA000045_23   2300001   2410000
BA000045_24   2400001   2510000
BA000045_25   2500001   2610000
BA000045_26   2600001   2710000
BA000045_27   2700001   2810000
BA000045_28   2800001   2910000
BA000045_29   2900001   3010000
BA000045_30   3000001   3110000
BA000045_31   3100001   3210000
BA000045_32   3200001   3310000
BA000045_33   3300001   3410000
BA000045_34   3400001   3510000
BA000045_35   3500001   3610000
BA000045_36   3600001   3710000
BA000045_37   3700001   3810000
BA000045_38   3800001   3910000
BA000045_39   3900001   4010000
BA000045_40   4000001   4110000
BA000045_41   4100001   4210000
BA000045_42   4200001   4310000
BA000045_43   4300001   4410000
BA000045_44   4400001   4510000
BA000045_45   4500001   4610000
BA000045_46   4600001   4659019
```

Continuation (36 of 47) of BA000045 from base 3500001 (BA000045 Gloeobacter violaceus PC

```
Query Match        27.7%;   Score 415.6;  DB 1;  Length 110000;
Best Local Similarity  57.7%;   Pred. No. 1.8e-60;
Matches  839;   Conservative  0;   Mismatches  0;   Indels  36;   Gaps  4;
```

```
RESULT 14
AP008231_05/c
WPCOMMENT
Sequence split into 27 fragments   LOCUS AP008231 Accession AP008231
Fragment Name     Begin       End
AP008231_00           1    110000
AP008231_01      100001    210000
AP008231_02      200001    310000
AP008231_03      300001    410000
AP008231_04      400001    510000
```

```
Db    ||||||||                       |||||| |    ||||||||  |||||||||| |||||||||||  |||
      86837  TGACAGTGCGGACGTTTGGGCCAATCCTCAGTTCTTTGCCCTCGATCCTGAAACGGGCGC  86778
Qy       723  CCCCACGGTGCGGCGGCCGGTTCCGGCGGCCGGGGGCCATTCCGCCGGCCCAGCTCTGGG  782
Db    ||||||||||  ||||| ||     ||    ||||||||||| ||||||||||||||||||
      86777  AGTTGATCAGACGAGGCCCGTGTGCCGCCCGACTATTTCTCCGAAACCGGCCAACTCTGGG  86718
Qy       783  CAATCCGCTCTATCGCTGGAGATGTGATGAAAGGACAACTTGCTCGGTGCATTGCCCG  842.
Db    |||||||||||| |||   ||||||||| |||||||||||||||||||||||||||||||
      86717  CAATCCCGTCTACAACTGGGCTGCGTCAGGCGGCCGATGGCTATCGCTGGTGTGCAACG  86658
Qy       843  CATAAGGCAGTCGCTCAAGCAGTGCCACCTGGTGCGTGCGGCATCGACCACTTCCGGGTTTGA  902
Db    ||||||| | |    ||||   ||||||||||||||||||||||
      86657  GCTGCAACAGCTCCTCAGCTTAGTGGACTACATTCGCATCGACCACTTCCGCGGTTTAGA  86598
Qy       903  AGCCTACTGGGAGGTTCCGTTTGGCCGGTCAAATGCTGGAGGGGCGCTGGGTCAAAGC  962
Db    ||||||||||  |||||  |||  |||||||  ||||||||||||||||||||||||||
      86597  GGCGGTTTGGTCGGTTCCCGTGGTGAAGAAACGGGGATCGACGGAGTGGGTCAAAGC  86538
Qy       963  CCCAGGGGGAGAAGCTGTTTGCTGCTGGACACGATTCGGCAAAAACTGGGAGGCGTGG  1022
Db    |||||| ||||||||| |||||||||||||||||||||||
      86537  CCCAGGCGCGCTGATCTGCTGAGCACGATTCGCAAAAACTGGGAGACGCGCTACCGATTCTGGC  86478
```

Continuation (6 of 27)

```
Qy      1023  CGAAGACCTGCGGGGTGATCACCCCGAGGTGGAGGCTTTGCGCGATGGCTTCGGGTTCCC  1082
Db    |||||||  |    |||| ||||||||||||||||||||||||||||||||||||||||||
      86477  AGAGGATCTCGGTGTGATTACGCCGGAGGTGGAAGCGCTGCGCCGATCGCTTTGAGCGTGCC  86418
Qy      1083  CGGCATGAAGATTTTGCAGTTGCTTTCCGTGGTGCTCGGGGATTGGGGGC  1142
Db    ||||||| |||| ||||||  |||||||||||  |||||||||||||||
      86417  GGGCCATGAAGATTCTGCAGTTCGCCTTGACTCGGGCCAATGCCAATGCCTATCTACCCGCA  86358
Qy      1143  CTACCCCGCGCACGGCAATGTGGTGTGTACAGCGAACCCACGAACCACCACCCT  1202
Db    ||||||||||  |  ||||| |||   ||||||||||||||||||
      86357  CAACTACTGGGGTCGTCGCTTACCAGCGCGGCCTTCATGCGGGCCTACCTGGCCCG  86298
Qy      1203  GGGATGGTTCCGCACCGGCCGAGCGAGCCGAGCCCTTCATGCGGGCCTACCTGGCCCG  1262
Db    |||||||  ||||  |||    ||||||||||||||||||||  |||||||||||
      86297  CGGCTGGTTCCTGTCCCGCAATGACAGCGATCGCCAAACGGTGCTCGGGGCGCTGGGCGC  86238
Qy      1263  CTATGCGCATCCGTTGTTTGTCGGAATACGAGGTCGCGGGCGCTTTGATCGAGCTGCCTT  1322
Db                                 |||||  ||||||  ||||||||||||||
      86237  AGA--------------GTCGGGGCTGGGAAATTGAGTGAGAGCTGATCCGCCTTGGCTTG  86193
Qy      1323  CAAAAGCCCGGCCAAGCTGGCTATTGTGCCTTGCAGAGGATGGCTGCTGGGCCCCGA  1382
Db    ||||||||||||||  |||||| |||||||||||||||||||||  |||||||||||||||
      86192  GAGCTGCAGAGATTGGGCGATCGCAACTCCTCAAGATGTCTTCGGGGTTGGATAGCAG  86133
Qy      1383  GGCCCGCCATGAAACTTCCCCCGACGGGGACAACTGGGCGTGGCCTACGCCGAAGG  1442
Db    |||||||||||||  |||||  ||||||||||||||||||||||  |||||||||||||
      86132  CGCCCCGCATGAATCGAATCGACCCGGGCAAGCCCACCGGGCGCTTCGAGTGCCGA  86073
Qy      1443  CGACCTCGAGCCCGCTCGCCGCGCGGGGCCCTGCCCGGAG  1488
Db    |||||||||| |||||| |||  ||||||||||||||
      86072  CTGGCTGACGCGGCGGATCGTGCCCAACGCGCCTTGCGGGGGCACTCTCGCAG  86027
```

RESULT 15
BA000022_02/c
WPCOMMENT
Sequence split into 36 fragments LOCUS BA000022 Accession BA000022

Fragment Name	Begin	End
BA000022_00	1	110000
BA000022_01	100001	210000
BA000022_02	200001	310000
BA000022_03	300001	410000
BA000022_04	400001	510000
BA000022_05	500001	610000
BA000022_06	600001	710000
BA000022_07	700001	810000
BA000022_08	800001	910000
BA000022_09	900001	1010000
BA000022_10	1000001	1110000
BA000022_11	1100001	1210000

```
AP008231_05    500001    610000
AP008231_06    600001    710000
AP008231_07    700001    810000
AP008231_08    800001    910000
AP008231_09    900001    1010000
AP008231_10    1000001   1110000
AP008231_11    1100001   1210000
AP008231_12    1200001   1310000
AP008231_13    1300001   1410000
AP008231_14    1400001   1510000
AP008231_15    1500001   1610000
AP008231_16    1600001   1710000
AP008231_17    1700001   1810000
AP008231_18    1800001   1910000
AP008231_19    1900001   2010000
AP008231_20    2000001   2110000
AP008231_21    2100001   2210000
AP008231_22    2200001   2310000
AP008231_23    2300001   2410000
AP008231_24    2400001   2510000
AP008231_25    2500001   2610000
AP008231_26    2600001   2696255
```

Continuation (6 of 27) of AP008231 from base 500001 (AP008231 Synechococcus elongatus PC

Query Match   26.8%;   Score 403.2; DB 1; Length 110000;
Best Local Similarity 56.1%;   Pred. No. 2.2e-58;
Matches 834;   Conservative 0; Mismatches 628;   Indels 24; Gaps 3;

```
Qy        12  ACGGCGTTTTGGAATTTGCTCACCCCACCAGTTTCCGGGTCGCTCGGGGGATTGGGGC  71
Db    ||||||||||  |||||||  ||  ||||||||||||||||||||||  |||||||||||
      87497  ACGAGCCGCGGCATTTTGTTACATCCCACTTCGGTCGGCCCATTGCGGGGCCAGCGGCGCA  87438
Qy        72  TCTGGGCCCGGAGGCCGAGCGGTTTTGGACTGGCTGGCTGGCCGATGCGGGAGCCCCGCTGGTG  131
Db    |||||||  ||||||||||||| |||||  ||||||||||||||||||||||||||||||
      87437  CCTTGGTCGGCCCTCGCGGCGCCAGTTTCTTGACTGGTTGGCAACGGCGGGACAACAACTGTG  87378
Qy       132  GCAGGTCTTACCGCTGGGCGTCCACCAGTTACGGCGCCGTACCGTTCCTTCTCCGGC  191
Db    |||||||  ||| |  ||||||||||||||||||||  ||||||||||||||||
      87377  GCAAGTGTTGCCCCCTTGGGCGCGACAGGCTATGGCTATTGGCCTATTCGCCTTACCTCTGCTATTCCGC  87318
Qy       192  TTTTGCCGGTAACCCGTATTTGGTTGACCCCGAGATGCTGATTGAAAAAGGCTGGCTGA  251
Db    |||||||||||||||||||||||||||||||||||||||  |||||||||||||||
      87317  CTTGGCTGGCAATCCGCTCTGATCAGCCCTGAACTCTTCGGCAGAGATGGCTGCTGGCTCCA  87258
Qy       252  ACAA------AGGCAGCGCCCCGCGTATCCGACCCCAGCGCGCGTGGATTTATGCTGGCT  305
Db    ||||      |||||||||||||||||||  ||||||||||||||||
      87257  AGAATCGGACACTGGGCAGCAGGCCCTGCGTCCTCGCTTCCGAGCGATCGCGTCGATTTTGCCAGCGT  87198
Qy       306  TTACCAGACCCGCTGGCCCCGTTGTGGCGGGGCTTTCGCGGGGTTTCGGGCCAAGGGCTTC  365
Db    ||| |||||  |||  ||||||||||||||||||||||||||||||||  ||||||||||||
      87197  CTTGCCTATCGCGAATCAACTGCTGCCGCCGGTCCGTGCGCCGTCGCCAGCCAAAGAGAGCGGC  87138
Qy       366  GGCCCAGGATAAGACCGACTGGAAGCCTTTATCGGAGGCCGAGCGCTTCGCCGCTGGGAAGA  425
Db    ||||||||||||||||||||||||||||||||  |||||||||||  ||||||||||
      87137  TTCCAGGATCGCCAACTCCTTTCAAGCTTTCGTTGAACAGAGGAAGCAGCCCATTGGCTGGATGA  87078
Qy       426  CTATGCGCTTCTTTATGGCCCTCAAGACCCCGGTTGACGGCAAGCCCTGAACAGAGTGGAG  485
Db    ||| |||                 |||||||||| ||||  ||||||||||||||||||
      87077  CTACGCCCTTCCGTTGTTCATGGCGACCGTGACCTGCCCTGCAGGGGCCCGTGAGGGAGCTGGCCGGA  87018
Qy       486  CCCCGAGCGGTGCGGACCGGTGAACCGCTGCCCTGCTGCCTGCTGGGCCTCGAGGAGCTGGCCGGA  545
Db    ||||||||||||||  |   |||||||||||||||||||  ||||||||||||||||
      87017  GGAAGCGGCTGCGTCAGCGGCGCAACCTCAAGCCTTCGCTCTAAAGCCCCGATCGCTCTGGGGCGGCG  86958
Qy       546  GGAGGTGGCCCCTTTACGAGTGGATTCAGTTGGCTTTTTTATCTGGAATGGGGCCCAGACCAA  605
Db    |||| ||  |||||||||||||  |||||||||||||||||||||||||||||
      86957  CGAAATTGCTTCCCAGCAGTTTCTGCAGTGGCAATTTCTGGAGCAGTGGTGTTCGGCCCTCTCG  86898
Qy       606  GGCCTATGCCGGAATCCAAGGGGATTCAGATTATCGGCGATATGCCCATCTTTGTGGGCCTT  665
Db    |||||||||  |||||||||||||||||||  |||||||||||||||||||||||||
      86897  GGAAGAAGCCCAAGCCCGCCATATTTCGCTGATTGGCGATATTCCGATATTCGCGGTCACGTCGTCA  86838
Qy       666  CGATTCCTCAGATGTCTGGCGGCCACCCGCCAGTACTTCTACCTGCGAGGCCGA---TGGCAA  722
```

```
BA000022_12    1200001   1310000
BA000022_13    1300001   1410000
BA000022_14    1400001   1510000
BA000022_15    1500001   1610000
BA000022_16    1600001   1710000
BA000022_17    1700001   1810000
BA000022_18    1800001   1910000
BA000022_19    1900001   2010000
BA000022_20    2000001   2110000
BA000022_21    2100001   2210000
BA000022_22    2200001   2310000
BA000022_23    2300001   2410000
BA000022_24    2400001   2510000
BA000022_25    2500001   2610000
BA000022_26    2600001   2710000
BA000022_27    2700001   2810000
BA000022_28    2800001   2910000
BA000022_29    2900001   3010000
BA000022_30    3000001   3110000
BA000022_31    3100001   3210000
BA000022_32    3200001   3310000
BA000022_33    3300001   3410000
BA000022_34    3400001   3510000
BA000022_35    3500001   3573470
```

Continuation (3 of 36) of BA000022 from base 200001 (BA000022 Synechocystis sp. PCC 6803

```
Query Match         21.9%;  Score 329.2;  DB 1;  Length 110000;
Best Local Similarity 53.3%;  Pred. No. 6.4e-46;
Matches  800;  Conservative  0;  Mismatches  673;  Indels  27;  Gaps  4;
```

```
Qy      1  ATGCAACTCCAACGCGCTTTTGGAATTTGCTCCACCCACCAGTTTCGCTCCTTGCCCACCGCCAGTTTTCCGGGTCGCTGG    60
              ||||||  ||  ||||    ||||||||||||    ||||||||    ||||||||||||||||      ||||  ||||
Db  47019  ATGTTAGATAAACGTGTAGCGGTATTCTGCTCCATCCCACCTCCGCTCCTTGCCCAGTCGTTTC                   46960

Qy     61  GGGATTGGGGCTCTGGGCCGGAGGCCGAGCGGTTGCTTTCAGTTCATCGATTTTTGGCGGACCGGACGGAGAC         120
              ||||||||  ||  ||||    ||||||  ||||||||||||||||||||  ||||||||||  ||  |||||||
Db  46959  GGCATTGGGACTTGGGGACGGTGCTTTACGGTCTTTCAGTTCATCGATCGATTTTTGGCGGCGGACTGCCGAGAC        46900

Qy    121  GCCCGCTGGTGTGGCAGTGTGTACCGGTCTTACGGCCCTACCAGTTACGGCGACTGCCGTACCAGG               180
              ||||||||    ||||  ||||||  ||||||||||||||||||||||||||  ||||||||  |||
Db  46899  CAGAGTGTATGGCAAATATTGCCCGTTGGGCCAATATTGCCTGGCCCCCACCGGCCCCCCACCGGAAT             46840

Qy    181  TCCTTCTCGGCTTTTGCCGGTAACCCGTATTGGTTGGCAGATGCTGATTGAAAAA              240
              |||||||  |||||||||  ||||||||||||||    ||||||||||||||||||
Db  46839  TGCTATTCTGCCCTAGCAATTAATCCTGGTTGGATTGATCCGCCTAGCTGCCTGAAAGAA         46780

Qy    241  GGCTGCTGGAACAAAG------CGAAGCGCCCCCGCGATCGCCAAG             294
              |||||||||||||||       |||||  ||||||  ||||||||
Db  46779  GGATTTTTGCCACCGAGTTTATTAGACACCAAGCGGCCTGGCCCCCCGTGTTG          46720

Qy    295  TATGGCTGGCTTTACCGCTGTGCCTGCGTGGCCTGGGTTTCG            354
              ||||||||||||||||||||||  ||||||||||||||||||
Db  46719  TATGACCAGGCGGATCGCCTGACAAATCCAAGTTTAAAACAGGCCTTTGCCCTTTGCCCGG     46660

Qy    355  GCAAGGGGCTTCGGCCCCAGAGATAAGACCCGGCGCTGGCCAAGCCCAAGTGAT             414
              |||||||||  ||||||||||||||||||  |||||||  ||||||||||||||
Db  46659  ACCAACATAGAGCTAGCAATTGAGCAGGAATTTGCAGAGAATTTGCCAAGCCCAAGTGAT         46600

Qy    415  TGGCTGGAAGACTATGCGTCTTTATGGCCCTCAAGACCCCGTTGACGGCAAGCCCTGG             474
              ||||||||||||||||||  |||||||||||||||||||  ||||||||||||||||||
Db  46599  TGGCTAGCAGATTACCGCCCGTTCATGGCCATCAAAGAGGCCCACATGGAGAGCCGGTTGG       46540

Qy    475  AACGACTGGAGCCCCCAGAGCGCCGACAGTTGCTCTGCGCGGGAACCCGGAAGCCGGAGAA              534
              ||||||  ||||||||||  |||||||||||||||||||||||  ||||||  |||||
Db  46539  CATCAATGGGACAAGGACAGTTTTATACCATCAATTTTTCGCAATTTGCAAATTCTTTGTG       46480

Qy    535  GAGCTGCCGCAGGAGTGCGGGCCCTTTACGAGTGGCTCGGCTTCTTTATCTGGAATGG         594
              ||||||||||||||||  |||||||  ||||||||||||||||  ||||||||||||||
Db  46479  CGCCTGAAAACGGAAGTTTTATACCATCAATTTTCTTGGTTTTCGCCGATATGCCCATC        46420

Qy    595  GGCCAGACAAGGCCTATGCCGAATCCAAGCCATGGGATTATCGGCGATGCCCATC        654
              ||||||||||  ||||||  ||||||||||||  ||||||||||||||||||||||
Db  46419  CAGGAAGTCAAAGCCTAACGACCACGGCACATAGCCGGACCTCGCCCATCGGGGGGAC     46360
```

```
Qy    655  TTTGTGGGCCTTGATTCCTCGGCCACCCGGCAGTACTTCTACCTCGAGGCC          714
              ||||||||||||||||    ||||||||  |||||||||    ||||||||
Db  46359  TACGTTGCCCACGATAGCGCTGACGTTTGGGCCAATCCAGAGAACTTTGGCCTTGATCCG    46300

Qy    715  GA---TGGCAACCCACGCGGTGGTGGCGGCGTTCGCGGGCGGTACCACCGGACCACAGGGC         771
              ||   ||||  ||||||||||||||||  ||  ||    ||||||||||||||  ||  |||
Db  46299  GAAACCGGAGAAGCGGCCATGATGCGCCGGGTACCACCGGACTATTTCAGTGCCACAGGGG     46240

Qy    772  CAGTTCTGGGGCAATCGCTCTATCGTCCTGGGATGTGATGGAAAGGGACAACTTTGCCTGG        831
              |||||||||||  ||||||||||  ||||||||||||||||||  ||||||||||||||||
Db  46239  CAACTCTGGGGTAATCCTGTGTACGACTGGGAAACCCTCAAAGCTACGGGCTTTGCCTGG      46180

Qy    832  TGCAATGCCGCATAAGGCAGTCGCTCAAGCAGTGCCACCTGGTGCGCATCGACCACTTC         891
              ||  ||  ||||  ||||||||  ||||  |||||||  ||  ||  ||||||||||||||
Db  46179  TGGATTAAGCGTTTTAAAGCCAATCTGCAATATCTAGACATTGTCCGCCATTGACCATTTC     46120

Qy    892  CGCGGGTTTGAAGCTCACTGGGAGGTTCCGTTTGCGCGGCCCAATGCTGTGTGAGGGGCGC        951
              ||  |||||||||||| ||||||  ||||| |||||||||||||||||||| | |||||||
Db  46119  CGGGGGCTTCGAGTCCTATTGGGGAGTGCCCAAGGGGAAAAAACTGCTGAAAATGGCGAA      46060

Qy    952  TGGGTCAAAGCCCACAGGGGAGAGAAGCTGTTTGCTGCGGGTGCGGGCCCAACTGAGCG---AT       1008
              ||||| ||||||||  ||||||||||||||||||  |||||||  ||||||||  ||||    ||
Db  46059  TGGGTATCCAGCCCCGGCAAGGAATTTTTCCAAGCCCTAGGGAAAGCCCTGGGTGATAAT       46000

Qy   1009  GCGGCCCCATCATTGCCGAAGACCTGGGGGTGATCACCCCGAGGTGAGGCTTTGCGCGGAT        1068
              |||||||||| |||||||||||  |||||||| ||  ||||||  ||||  ||||| ||||
Db  45999  TTACCCATTGTGGCCGGAAGATTTGGGGGTAATTACTCCGGAAGCGGGAAGCGCTACGGAT      45940

Qy   1069  GGCTTCGGGTTCCCGGGCATGAAGATTTGCAGTTTGCATTTCCATTTCGCCTTTGACTTCGCCGACCGGGGTAAT      1128
              |||||  |||||||||||||||| |||  || ||||| ||||||||||| ||||| ||  |||      |||||
Db  45939  GAATTTAACTTTCCCGGCATGAAATGTCTCCATTTCGCTTTTCCCCTCCAGTACTCTC                   45880

Qy   1129  TTTTTGCCCACAACTACCCCGCACGGCAATGTGGTTGTACAGCGGAACCCACGAC           1188
              ||||||||| |||||||  |||||||||||||  |||  ||||||  ||||||||
Db  45879  CCCTTTCTGCCCTTCAACTACCAGCAATGGCAACGCGCGGTGGTTTATACCGGACACCCACGAC    45820

Qy   1189  AACGACCACACCCTGGATGGTTCCGCACGGGCGGCCGAGCCGGGCCCTCATGCGG--------       1248
              |||||| || ||| ||| |||  ||||  |||||| ||||||    ||||||| ||
Db  45819  AACGACACACCACCGTGGGCGTGGCTGGTTTCAGGAACCGTCAGGAGGATGATCAGC-            45771

Qy   1249  GCCTACCTGGCCCGCTATGGCACCGGGCTGTGTCATGGGATTGTTGTGCGAAATACGAGGTGCGGCGCTTG     1308
              |||||||||||||||||||||  |||||||||| |||||  ||||||||||||||||||||||  ||  || |
Db  45770  ----AAAGGTGATCAATTACCTCGGCTGTGTGTGGCGCAACGAAGGTATCCATTGGAGACGTGCTG        45715

Qy   1309  ATCGAGCTGGCCTTCAAAAGCCCGGCCTATCCGAAGCGCGCCCAGCGCGTGGAT        1368
              ||  ||||||||||||||||||||||||| |||||||   |||  ||||||||
Db  45714  ATTCGCCTTAGCCGTCCCAGTTCGGTGGGGCGGGGTCCCCCTCCAGGAGCATCCTC       45655

Qy   1369  GGGCTGGGCCCCGAGCGCCCGATGAACTTCCCCGGACGACTTCGCGGGGTTTCG         1428
              |||||||| | ||| ||||||||||||| |||||| |||  |||||   ||| |
Db  45654  GGTTTGGGTAGTGACTGTCGCATGAACTTACCCGGGCCAACTGGGGTTGG              45595

Qy   1429  CGCTACGGCCGAAGGCGACCTCGGCATCGGGGCATCGGGGGCCGCCCGCAG        1488
              |||||| |  ||||||||||||||||||  ||||| |||   ||  | ||
Db  45594  CGTTACCATCCCGACCAACTCAACGATTGGCTATCGGGGCATTTGAGTTTCATTACGGAG     45535
```

```
Search completed: January 14, 2006, 13:11:42
Job time : 7432 secs
```

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    January 14, 2006, 10:18:26 ; Search time 5555 Seconds
                                (without alignments)
                          12659.043 Million cell updates/sec

Title:         US-10-618-976-1
Perfect score: 1503
Sequence:      1 atgcaacccaacgcgcttt..........ccgaggccagccagcgcgct 1503

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:      41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters:    82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

Database :   EST:*
```
               1:  gb_est1:*
               2:  gb_est2:*
               3:  gb_est3:*
               4:  gb_htc:*
               5:  gb_est4:*
               6:  gb_est5:*
               7:  gb_est6:*
               8:  gb_est7:*
               9:  gb_gss1:*
              10:  gb_gss2:*
              11:  gb_gss3:*
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	
1	156.6	10.4	1569	10	CL973535	CL973535 OsIFCC024	
2	149	9.9	877	10	CG848343	CG848343 ZMMBBb031	
3	145.4	9.7	719	7	CK369120	CK369120 zmrws055_	
4	145	9.6	1875	4	CNS0A259	BX829726 Arabidops_	
5	143	9.5	762	7	CV470446	CV470446 43319.1 C	
6	140.2	9.3	733	1	AJ795827	AJ795827 AJ795827	
7	138.2	9.2	538	3	BI720945	BI720945 1031053A0	
8	132.4	8.8	1155	4	AY109140	AY109140 Zea mays	
9	128.2	8.5	678	7	CK767135	CK767135 pam01-9ms	
10	126.6	8.4	683	6	CB074139	CB074139 EST01770	
11	125.4	8.3	646	7	CO119845	CO119845 GR_Eb022	
c 12	120	8.0	605	2	BE515486	BE515486 WHE0613_B	
13	117.8	7.8	548	6	CF013402	CF013402 QBK5e12.x	
14	117.4	7.8	765	6	CF437795	CF437795 EST674140	
15	116.6	7.8	769	7	CV302316	CV302316 67345.1 S	
16	114.2	7.6	576	3	BJ175019	BJ175019 BJ175019	
c 17	114.2	7.6	577	3	BJ184679	BJ184679 BJ184679	
18	112.6	7.5	666	6	CA067878	CA067878 SCQSAD105	
19	112.4	7.5	609	6	CD481918	CD481918 atr01-9ms	
20	110.4	7.3	403	3	BJ943583	BJ943583 BJ943583	
21	108.8	7.2	518	1	AW623584	AW623584 EST321529	
22	107.8	7.2	832	7	CO094503	CO094503 GR_Ea16K	

23	107.2	7.1	471	6	CB210624	CB210624 OML00904
24	106.6	7.1	606	6	CD442683	CD442683 EL01N0414
25	105.2	7.0	706	3	BG600511	BG600511 EST505406
26	103	6.9	662	3	BJ940916	BJ940916 BJ940916
27	102.2	6.8	607	8	CX189158	CX189158 77-E02303
28	92.8	6.2	546	6	CA797031	CA797031 Cac_BL_41
29	91.2	6.1	508	3	BI946157	BI946157 su78f06.y
c 30	87.4	5.8	301	3	BM363539	BM363539 BS320057A
31	86.4	5.7	802	6	BJ977741	BJ977741 BJ977741
c 32	86.2	5.7	716	6	CF512764	CF512764 CAbud0004
33	84	5.6	582	6	CD482306	CD482306 atr01-29m
c 34	83	5.5	395	7	CO984423	CO984423 GM89013B1
35	81.2	5.4	626	7	CK756616	CK756616 atr02-5ms
36	78.6	5.2	800	1	AI823034	AI823034 L30-888T3
37	78.2	5.2	890	8	BE055692	BE055692 GA_Ea003
38	77.8	5.2	695	7	CK718248	CK718248 18239 Swo
39	77.8	5.1	714	3	BI405594	BI405594 140H11 Ma
40	76.8	5.1	636	2	BG858024	BG858024 102405SD0
41	74.4	5.0	524	1	AI725928	AI725928 BNLGHI136
42	74	4.9	744	9	CC943404	CC943404 BOICU30TR
43	70.8	4.7	887	2	CO127810	CO127810 GR_Eb12J
44	69.6	4.6	513	2	BF004941	BF004941 EST433376
45	66.2	4.4	314	6	CA999774	CA999774 S234J_A07

ALIGNMENTS

```
RESULT 1
CL973535
LOCUS          CL973535                    1569 bp    DNA    linear    GSS 21-SEP-2004
DEFINITION     OsIFCC024465 Oryza sativa Express Library Oryza sativa (indica
               cultivar-group) genomic, genomic survey sequence.
ACCESSION      CL973535
VERSION        CL973535.1  GI:52401596
KEYWORDS       GSS.
SOURCE         Oryza sativa (indica cultivar-group)
ORGANISM       Oryza sativa (indica cultivar-group)
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
               Ehrhartoideae; Oryzeae; Oryza.
REFERENCE      1  (bases 1 to 1569)
AUTHORS        Ma,L., Wang,C., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
               Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
               Wong,G.K.S., Deng,X.W. and Wang,J.
TITLE          An analysis of transcriptional regulation of the rice genome and
               its comparison to Arabidopsis
JOURNAL        Unpublished (2004)
COMMENT        Contact: Chen Chen
               Department of Bioinformatic
               Beijing Institute of Genomics
               Chinese Academy of Sciences, Beijing 101300, China
               Tel: 86-10-80481559
               Fax: 86-10-80488676
               Email: chenchen@genomics.org.cn
               Rice genomic sequence.
               Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..1569
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

Query Match             10.4%;    Score 156.6;  DB 10;  Length 1569;
Best Local Similarity   50.9%;    Pred. No. 6.8e-28;
Matches  398;  Conservative   0; Mismatches 381;  Indels   3;  Gaps   1;

Qy     415 TGGCTGGAAGACTATGCGCTCTTTATGGCCCTCAAGCACCCGGTTTGACGGCAAGCCCTGG 474
           ||||||||||||||| |||| ||||||||||||||||||  ||| ||| | |||||||||||
Db     532 TGGCTTGAAGATGCTGCACTTTTGCTGCATTGACAGAAGTATTGATGCATTATCCTGG 591
```

us-10-618-976-1.rst

Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: rwing@genome.arizona.edu
PCR PRimers
FORWARD: T7
BACKWARD: M13r
Plate: 0319 row: P column: 13
Seq primer: M13r
Class: BAC ends.

FEATURES   Location/Qualifiers
    source   1..877
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBb0319P13"
                     /lab_host="DH10B"
                     /clone_lib="ZMMBb"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; Zea mays L. ssp. mays"

ORIGIN

Query Match   9.9%; Score 149; DB 10; Length 877;
Best Local Similarity 52.5%; Pred. No. 4.9e-26;
Matches 351; Conservative   0; Mismatches 315; Indels 3; Gaps 1;

Qy   571 CAGTGGCTTTTTATCTGGAATGGGGCCAGACCAAGGCTATGCGAATCCAAGGGGATT 630
               |||||| ||||||||||| |||| || | ||| || || | | |||||||
Db   78 CAATTCCTGTTCCAAAGGCAGTGGAAGAAAATTCGGCACTATGCACAGACAAAGGGAGTC 137

Qy   631 CAGATTATCGGCGATATGCCCATCTTTGTGGCCTTCTGAGATGTCTGGGCCAAC 690
               | |||| ||||| |||| ||| ||| || ||| |||| || | ||||
Db   138 AGCATAATGGGAGACATGCCTATTTATGTAGGTTATCACAGTGCTGTGATGTTTGGGCAAAT 197

Qy   691 CCCAGTACTTCTACCTCGAGGCGATGGCAACCCCACGGTGGCGGCGTTCCGCGG 750
               |||| || | |||| | ||| ||| || ||| |||||| | |||||||||
Db   198 AAGAAACATTTTTGCTGAATAGGCATGGTTATCCTATGCTAGTCAGTGTGTTCCTCCT 257

Qy   751 GACTACTTCTCCGAAACCGGCCAGCTCGGGGCAATCCGCTGGGGATGTGATG 810
               || ||||| || || | ||| || ||| ||| | || |||||||
Db   258 GATGCCTTCAGTGATACTGGTCAGCTCGGGGAGTCCTCTATATGATGATGTGGAAAGCCATG 317

Qy   811 GAAAGGGACAACTTGCCTGGTGCATTGCCCGGTACGGCAGTCGCTCAAGCAGTGCCAC 870
               || | ||||| |||| || | || | |||| ||| | || ||| ||| ||| |||
Db   318 GAGAAAGATGATTTTTATGGTGATAAGTCGCATTAGACGTGCACAGACTCTATATGAT 377

Qy   871 CTGGTGCCATCGACCACTTCCGGCGTTTGGCCGGTTCCGTTTGGCCGG 930
               |||||||| ||||| | | |||| | ||||| |||||||| ||||||
Db   378 GAATTCAGGATAGATCACTTTAGAGGATTGCTGGCTTTTGGGCTGTCCCTTCTGACAAGCG 437

Qy   931 CCCAATGCTGTGTGAGGGGGGCGCTCGGGTCAAAGCCCCAGGGGGAGAAGCTGTTTGCTGCGCGTG 990
               ||| | | | ||| || | || || ||||| | | ||| |||| | || |||||| ||| ||
Db   438 AAAGTTGCAATGGTTGGACGATGGCGCGCCCATCAGTTGGTCCTCAAAAATCTTTGTTCGACGCCATC 497

Qy   991 CGGGCCCAACTGAGCGGATGCGCGCCCATCAGTTGCGAAGACCTGGGCGTGATCACCCCGAG 1050
               |||||| | | ||| | ||| |||| ||||| ||| |||| || | ||||| | ||
Db   498 TCCAAAGCAGTTGAAAGATTAATATCATAGCTGGAGTTATAACTGAGAT 557

Qy 1051 GTGGAGGCTTTGCGCGATGGCTTCGGGTTCCCCCGGCATGAAGATTTGCAGTTTGCTTTT 1110
               ||||| | ||| |||| |||| ||||||| || || ||| ||||||||| || |||||
Db   558 GTAGTTCAGCTCAGGAGAAATATATTGGAGCACCTGGAATGGCTGTTTCTCCAGTTTGGTTTT 617

Qy 1111 TCCGGTGAGGACAACGCCCTTTTGCCCACAACTACCCCGCAATGTGGTGGTG 1170
               |||| || | ||||| | ||||| ||| || | |||| | | ||||| ||
Db   618 GGTAGTGACTCTAGCAACCCTCATTTACCTCACAATCACCAGCA---CAATCAAGTTGTG 674

Qy 1171 TACAGCGAACCCACGACAACGACACCACCTGGGATGGTTCCGCACCGAGCGGGCC 1230
               || |||| | ||| || ||| ||| ||| | |||||||| ||| ||| ||||
Db   675 TACACTGGAACACATGACAATGACGATCCAGGATCCAAGGTTGGTGGAGCAATTTGATGGAACAA 734

Qy   475 AACGAGTGGAGCCCGAGCTGCGCGACCGTGCGCTGCCCTGGCCAGGGCCCGTGAG 534
               ||||||
Db   592 TATGAGTGGCCTGGCCTGGAACCCTCGAAAAACCGACATCTCCGGGGCCTTGGAAGACATATATCAA 651

Qy   535 GAGCTGCCGCGACGAGGAGGTGGCCCCTTTACGAGTGGATCAGTGGTCTTTTTATCTGGAATGG 594
               | ||| || |||
Db   652 AAACAGAAGGATTTCATAGAGATATTTATGGCTCAACAATTCTTATTTCAAAGGCAGTGG 711

Qy   595 GGCCAGACCAAGGCCTATGCCGAATCCAAGGGGATTCGGGCGATATGCCCATC 654
               || ||||| |||||| | || ||| ||||||||
Db   712 CAGCGAATTCGTAAATATGCAAAGAAGTTGGGTATCAGCATAATGGGTGACATGCCCATA 771

Qy   655 TTTGTGGCCTTCGATTCCTCAGATGTCTGGGCGCCAACCCGCAGTACTTCTACCTCGAGGCC 714
               |||| |||| || || | | || ||| ||||| | ||||| | | || || | ||||
Db   772 TATGTTGCTACCATAGTCGCAGATGTTGGGCAAACAGGAAATCGTTTTGTGTTCGTTGGACAAG 831

Qy   715 GATGGCAACCCACGGTGTGTGCGGCGTTCCGCGGGACTACTTCTCCGAAAACCGGCCAG 774
               |||| || ||| ||| | | || ||||| | |||| ||| || ||| || |||||||
Db   832 AATGGTTTTCCGACTTTTGTTAGTGGTGTTCCACCTGATGCATTTAGTGAAACTGGTCAG 891

Qy   775 CTCTGGGGCAATCCGCTCTATCCGCTGGTGCGCCATCGGTCGCTCAAGCAGTCGCTCACTTCCGC 834
               || |||| ||||| || | ||| | || |||| || ||| ||| ||| || |||
Db   892 CTTTGGAACAGTCCATTGTACGACTGAAAGCTATGGAAGCAGGTGATTTGAATGGTGG 951

Qy   835 ATTGCCCCGCATAAGGCAGTCGCTCAAGCAGTGGCCACCTGGTGCGCCATCGACCACTTCCGGC 894
               || |||| ||||| | || | ||| || || |||||
Db   952 ATAAAAAGGATTAATCGTGCCCTGATTTGTATGATGCATCGTTCCGTATAGACCCATTTCCGG 1011

Qy   895 GGGTTTGAAGCCTACTGGAGGTTCCGTTGCCGGCCAATGTCTGTTGGAAGGGGCGCTGG 954
               |||||| ||||| || || | ||| | || || ||| | || ||| ||||||
Db 1012 GGGCTTGCAGGTTTTTGGGCAGTTCCTTCGAGTCAAAAGTACGCGCTCGTTGGAAGCTGG 1071

Qy   955 GTCAAAGCCCCCAGGGGGAGAAGCTGTGTTCGTGTTTGCTGTCGAGCGGATGCGCCCC 1014
               || |||| || ||||| || | | || | || ||| |||| || ||| ||| |||||
Db 1072 AGGGCTGACCGAGGAATGCATTTTTATGATGCATTATTCAAAGCTGTTCGGTAGAATAAAC 1131

Qy 1015 ATCATTGCCGAAGACCTGGGGGTGATCACCCCCCGAGGTGGAGGCTTTGCGCGATGGCTTC 1074
               |||||| ||||| ||||| ||| | || ||||||| |||||
Db 1132 ATAATAGACAGAAGACCTGGTGATAACAGAAGATGTTGTTGACCTAAGGAAGTCTATT 1191

Qy 1075 GGGTTCCCCGGCGATGAAGATTTTGCAGTTTGCTTTTTCCGGTGAGGACAACGCCCTTTTG 1134
               ||||||||| || ||||| ||||||| |||| | | || | |||| ||||||
Db 1192 GAGGCTCCTGGAATGGCAGTTTCTCAGTTTGCTTTTTGGTGGTCGTTCTGACAATCCACAC 1251

Qy 1135 CCCCACAACTACCCCGGCCGCAATGTGGTCGTGTACAGCGAACCCACGACAACGAC 1194
               ||||| |||||| ||||||| ||| | ||| |||| || |||| || || |||||
Db 1252 TTGCCCCATAAACCATG---AATTTGATCAAGTAGTATACACCCGGAACACAGACAATGAC 1308

Qy 1195 AC 1196
           ||
Db 1309 AC 1310

RESULT 2
CG848343
LOCUS   CG848343   877 bp   DNA   linear   GSS 18-NOV-2003
DEFINITION ZMMBb0319P13.r ZMMBb Zea mays genomic clone ZMMBb0319P13 3',
            genomic survey sequence.
ACCESSION   CG848343
VERSION   CG848343.1 GI:38375204
KEYWORDS   GSS.
SOURCE   Zea mays
  ORGANISM Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 877)
  AUTHORS   Yu,Y., Kim,H.R., Hatfield,J., Soderlund,C., Bharti,A.K., Messing,J.
            and Wing,R.
  TITLE   Sequencing of the maize genome
  JOURNAL   Unpublished (2003)
COMMENT   Contact: Rod Wing

Qy   1231 GAGCGGGCC 1239
           ||| ||| ||
Db   735 GAGAGGTCC 743

RESULT 3
CK369120
LOCUS   CK369120   719 bp   mRNA   linear   EST 23-DEC-2003
DEFINITION zmrws055_0B21-004-e09.s0 zmrws055 Zea mays cDNA 5', mRNA sequence.
ACCESSION   CK369120
VERSION   CK369120.1 GI:40335050
KEYWORDS   EST.
SOURCE   Zea mays
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 719)
AUTHORS   Bohnert,H., Sharp,R.E., Springer,G.K., Poroyko,V., Fredricksen,M.,
            Sharp,L.G., Spollen,W.G., Ries,J., Guillen,A., Khambati,A.,
            Topinka,C., Davis,G.E., Schachtman,D., Wu,Y. and Nguyen,H.T.
TITLE   NSF Grant DBI-021842: Functional Genomics of Root Growth and Root
            Signaling Under Drought
JOURNAL   Unpublished (2003)
COMMENT   Contact: Hans Bohnert
            University of Illinois, Urbana-Champaign
            1201 West Gregory Drive, Urbana, IL 61801, USA
            Tel: 217-265-5475
            Fax: 217-333-5574
            Email: bohnert@life.uiuc.edu
            POLYA=No.
FEATURES   Location/Qualifiers
     source   1..719
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="taxon:4577"
                     /clone_lib="zmrws055"
                     /note="   Library zmrws055 consists of the same cDNA
                     material as library zmrws05 (described below) but was
                     sequenced from the 5 prime end. The sequence identifier
                     uses the '.s0' suffix because the library tag was at the 3
                     prime end and thus not identified. Samples were
                     collected in Robert E. Sharp's lab (University of
                     Missouri-Columbia) to construct three normalized cDNA
                     libraries. Dark-grown maize seedlings with primary roots
                     12-20 mm in length were transplanted to high (-0.03 MPa)
                     or low water potential (-1.6 MPa) vermiculite, and
                     harvested at 5 h and 48 h after transplanting. About
                     1,000 roots were used for each of the low water potential
                     libraries (zmrws05 and zmrws48) while 500 roots were
                     combined from each of the two time points at high water
                     potential (zmrww00). Each root was divided into 4
                     segments (distances are from the junction of the root apex
                     and root cap): segment 1, 0-3 mm plus the root cap;
                     segment 2, 3-7 mm; segment 3, 7-12 mm; segment 4, 12-20
                     mm. (For details of conditions see (1) with nutrient
                     modifications as in (2)).   The three normalized cDNA
                     libraries were constructed in the lab of Hans Bohnert
                     (University of Illinois-UC). Total RNA was extracted by
                     the 'hot Phenol' method (Plant Molecular Biology manual.
                     D5: 1-13, 2nd ed., 1997). This method worked in
                     eliminating carbohydrate material present in the root
                     tips. The integrity of the RNA was verified by denaturing
                     agarose gels and spectrophotometry (ratio A260/280).
                     Poly(A)+mRNA was isolated twice from total RNA using the
                     Oligotex Direct mRNA kit (Qiagen). Poly(A)+ mRNA was
                     converted to double-stranded cDNA and tagged by using
                     modified Oligo(dT) primers. One of 4 sequence tags
                     corresponding to a different segment of the root was added
                     to the 3'-end of the modified Oligo(dT) primers, including
                     a NotI site and used to reverse transcribe the
                     segment-specific mRNAs into cDNAs. Each library contains

all four tags. A suffix (s1, s2, s3, or s4) has been
added to each sequence identifier to designate which
region of the root (Root_segment_1, 2, 3, or 4) the
sequence was found in based on the identification of the
tag. A suffix of s0 indicates that the sequence tag, and
hence the source segment, could not be identified. The
double stranded cDNAs were size-selected (>450 bp). Size
selected cDNAs were adaptored with EcoRI adaptors at both
ends, and then digested with NotI. The cDNA was
directionally cloned into EcoRI-NotI digested pBS II SK(+)
phagemid vector (Stratagene) and electroporated into
E.coli DH10B. The total number of white colony forming
units (cfu) in the primary libraries before amplification
was as follows: zmrws05: 3.37 x106; zmrws48: 4.87 x 106;
zmrww00: 3 x 106. The background of empty clones was less
than 1%. Inserts ranged from ~0.5kb to >2.5 kb, as
determined by PCR. Plasmid DNA from the primary
libraries then was converted to single-stranded circles
and used as a template for PCR amplification using the T7
and T3 priming sites that flank the cloned cDNA inserts.
The purified PCR products, representing the entire cDNA
population cloned in each library, were used as a driver
for normalization. Hybridization between the
single-stranded library and the PCR products was carried
out for 44 hours at 30C. Non-hybridized single-stranded
DNA circles were separated from hybridized DNA rendered
partially double-stranded and electroporated into DH10B.
The total number of clones with insert was: zmrws05:
2.0x107; zmrws48: 4.2x107; zmrww00: 1.1x107. The
background of empty clones was less than 2%. Insert size,
determined by PCR of the entire library, ranged from 0.5kb
to 2.5kb. (1) Sharp R E; Silk W K; Hsiao T C. Growth of
the Maize Primary Root at Low Water Potentials I. Spatial
Distribution of Expansive Growth. Plant Physiology
(Rockville). 87(1). 1988. 50-57.   (2) Spollen W G;
LeNoble M E; Samuels T D; Bernstein N; Sharp R E. Abscisic
acid accumulation maintains maize primary root elongation
at low water potentials by restricting ethylene
production. Plant Physiology (Rockville). 122(3). March,
2000. 967-976.
TAG_TISSUE=Not found
TAG_SEQ=Not found"

ORIGIN

Query Match   9.7%;   Score 145.4; DB 7; Length 719;
Best Local Similarity 52.3%;   Pred. No. 3.7e-25;
Matches 346; Conservative 0; Mismatches 313; Indels 3; Gaps 1;

Qy   571 CAGTGGCTTTTATCTGAATGGGCCCAGACCAAGGCCTATGCGAATCCAAGGGGATT 630
          ||| ||| ||| | |||||| || || ||||||||||||
Db   37 CAGTTCTTATTTCAAAGGCAATGGCAGCCGGATTCGTAAATATGCACAAAAGCTGGGTATC 96

Qy   631 CAGATTATCGGCGCGATATGCCCATCTTTGTGGCCTTCGATTCCTCAGATGTCTCGGGCCAAC 690
          ||| | ||| ||||| |||||| | ||||| || | ||||||||||||||||||||||
Db   97 AGCATCATGGGTGACATGCCTATATATGTTGGCTACCATAGCGCAGATGTTGGGGGAAC 156

Qy   691 CCGCAGTACTTCTCACCTCGAGGCCGATGGCAACCCCACCGGTGGTGGCGGCGTTCCGCGG 750
          |||||| | || ||||| ||||||||||| |||| | || | |||||||||||
Db   157 AGGAAATCATTTTTGCTGGACAAAAACGGTTCCCCACTTTCGTTAGTGGCGTTCCACCT 216

Qy   751 GACTACTTCTCCGAAACCGGCCAGCTCTGGGCAATCCGCTCATCCGCTCATCGCTGGGATGTGATG 810
          ||| |||||||| | || || | | | ||| | ||||||| ||||||||
Db   217 GATGCATTTAGTGAAACGGGGTCAACTATGGAACAGTCCATTGTACGACTGGAAAGCTATG 276

Qy   811 GAAAGGGACAAACTTGCCTGGTGCATTGGCGCATTGCCGCGGATGCGCTCAAGCAGTGCCAC 870
          ||| |||||| ||||||||||| | || |||| ||||||||||||
Db   277 GAAGCAGATTGTTTTTCATGGTGGATAAAGAGGATTAAACGTGCCCTTGATTCGTATGAT 336

Qy   871 CTGGTGCCGCATCGACCACTTCCGCGGGTTTGAAGCCTACTGGAGAGTTCCGTTTGGCCGG 930
          || ||| |||||||||| |||| ||||||||||| | ||| || | ||||| ||||||
Db   337 GAATTCCGTATTGACCATTTCCTCGGGGCTTGCGGGTTTTTGGGCAGTCCCTTCTGATGCA 396

Qy   931 CCCAATGCTGTGGAGGGGCGGCTGGTCAAGCCCCAGGGGAGAAGCTGTTTGCTGCGGTG 990

```
gene        /plasmid="pCMVSPORT 6"
            complement(1..1875)
            /gene="At5g64860"

ORIGIN

Query Match              9.6%;  Score 145;  DB 4;  Length 1875;
Best Local Similarity   50.6%;  Pred. No. 5.5e-25;
Matches 403;  Conservative 0;  Mismatches 390;  Indels 4;  Gaps 2;

Qy    415 TGGCTGGGAAGACTATGCGGCTCTTTATGGCCCTCAAGACCCGGTTGACGGCAAGCCCTGG 474
              |||||  ||     ||  ||      ||
Db    688 TGGCTTGAAGATGCTGCTTTATTTGCAGTCATTGACAATACTTTAAATGCATACACAGTTGG 747

Qy    475 AACGAGTGGAGCCCCGAGCTGCGGCGACCGTGCCCTGCCCGGCTGAACCGGCTCCAGGCCCCGTGAG 534
              |||||||                                           |||
Db    748 TTTGAGTGGCTGAACCACCACTAAAAACCGTCACTCTTTCTGCCTTGGAAGCTATATATGAA 807

Qy    535 GAGCTGGCCGAGGAGGTGGCCCTTTACGAGTGGATTCAGTGGCTTTTTTATCTGGAATGG 594
          |||                                           |||
Db    808 AGCCAAAAGGAGTTTATAGACTTGTTCATTGCTAAGCAATTTTGTTCCAAAGGCAGTGG 867

Qy    595 GGCCAGACCAAGGCCTATGCCGAATCCAAGGGGATTCAGATTATCGGCGATATGCCCATC 654
          |||||                   ||||||            |||||||||||||||||
Db    868 CAGAAAGTTCGTGAGTATGCACGGCGCAAGGAGTTGATATAATGGGAGATATGCCCATT 927

Qy    655 TTTGTGGCCTTCGATTCCTCAGATGTCGGGCAACCCGCAGTACTTCTACCTCGAGCCC 714
          |||||     |                          |||          ||
Db    928 TATGTAGGGATATCACAGTGCACAGTTGGGCAAATAAGAAACATTTCTTACTGAACAAG 987

Qy    715 GATGGCAACCCACGGTGGTCGGCGGGGTTCCGGGCGTTCCTCCGACTTGTTCAGTGAAACTGGTCAA 774
          ||     ||                    |||     |||||||  |||
Db    988 AAAGGCTTTCTCCTCTTCCTTGGTTCCGGTGTTCCTCTGACTTTGCTGCTGTGC 1047

Qy    775 CTCTCGGGCAATCCGCTCTATCGCTGGGATGCTCTGGGATGTGATGGAAAGGGACAACTTTGCCTGTGC 834
          ||  |||||| |||    |||      |||          |||   ||
Db   1048 CTGTGGGCAGCCCTCTTTATGACTGGAAAGCAATGAGAGTGACCAGTGCCATCGACCACTTCCGC 1107

Qy    835 ATTGCCCGCATAAGGCAGTGCTGCTCAAGCAGTCCACCTGGTCGCGCATCGAATGCAGGATTGATCACTTCAGA 894
          |||       ||||||                         |||       |||||  |||||||||||||
Db   1108 GTTAATCGAATAAGACGCGCACAGGACTTGTATGCAGCAGGATTGATCACTTCAGA 1167

Qy    895 GGGTTTGAAGCCTACTGGGAGGTTCCGTTGGCCGGCGCAATGGTGGGAGGGCGCTGG 954
          ||||||||||||||     |||    |||   |||        |||
Db   1168 GGATTTGCAGGGTTTGGGCCGGTCCCTTCTGAAGCGAAAGTTGCCATGGTTGCACGATGG 1227

Qy    955 GTCAAAGCCCCAGGGGAGAAGCTGTTTGCTGCGGCGTGCCGGATGCGCCCC 1014
          |||||       ||        ||| |||        ||    ||
Db   1228 AAGGTAGGAACCTGGAAAGTCATTATTTGATGCCATTTTTAAAGGCGTTGGGAAGATCAA 1287

Qy   1015 -ATCATTGCCGAAGACCTGGGGGTGATCACCCCCGAGGTGGAGGCTTTGCGCGATGGCGTT 1073
           |||| |  ||||||   ||                 |||      |||
Db   1288 AATCATAGCTGAAGATTTGGGAGTTATTACTAAAGATGTAGTTGAGCGTGAGGAAATCTAT 1347

Qy   1074 CGGGTTCCCCGGCATGAAGAGATTTGCAGTTGCTTTTTTCCGGTGAGGACAACGCCGCTTTT 1133
          |||||||||   ||  |||||| |||          ||||||            |||||||
Db   1348 CGGAGCACCTGGAATGGCCCGGTCCTCCCAATTTGCTTTTTGGAGGAGGCGCCGATAACCCACA 1407

Qy   1134 GCCCCACAACTACCCCGCGCACGGCAATGTGGTGTACAGCGGAACCCACCACGACGA 1193
          ||||       |||             |||
Db   1408 TTTACCTCACATCATG--AAGTAAACCAAGTTGTTTCCTCCTCGTGTACTCATGACAACGA 1464

Qy   1194 CACCACCCTGGGATGGT 1210
          ||   ||||||||
Db   1465 CACTATTCGAGGCTGGT 1481
```

RESULT 4

```
CNS0A259

LOCUS       CNS0A259           1875 bp   mRNA   linear   HTC  06-FEB-2004
DEFINITION  Arabidopsis thaliana Full-length cDNA Complete sequence from clone
            GSLTFB31F05 of Flowers and buds of strain col-0 of Arabidopsis
            thaliana (thale cress).
ACCESSION   BX829726
VERSION     BX829726.1  GI:42459525
KEYWORDS    HTC; GSLT cDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1875)
  AUTHORS   Castelli,V., Aury,J.M., Jaillon,O., Wincker,P., Clepet,C.,
            Menard,M., Cruaud,C., Quetier,F., Scarpelli,C., Schachter,V.,
            Temple,G., Caboche,M., Weissenbach,J. and Salanoubat.M.
  TITLE     Whole Genome Sequence Comparisons and 'Full-Length' cDNA Sequences:
            A Combined Approach to Evaluate and Improve Arabidopsis Genome
            Annotation
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1875)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     The sequences are based on single pass reads.
            Life Technologies (a division of Invitrogen) members carried out
            full-length librairies construction : Temple G.
            Genoscope members carried out sequencing and annotation : Castelli
            V., Aury J.M., Jaillon O., Wincker P., Menard M., Cruaud C.,
            Schachter V., Weissenbach J., Salanoubat M.
            URGV INRA : Clepet C., Caboche M.
            Annotation is based on the June 2003 version of the Arabidopsis
            genome released by MIPS (Munich Information center for Protein
            Sequences).5 prime and 3 prime are assembled with Phrap.
            http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/Full
            length
            http://www.genoscope.cns.fr/cgi-bin/ggb/ggb.pl?source=Arabidopsis.
FEATURES             Location/Qualifiers
     source          1..1875
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="GSLTFB31F05"
                     /tissue_type="Flowers and buds"
                     /ecotype="Col-0"
```

RESULT 5

```
CV470446

LOCUS       CV470446            762 bp   mRNA   linear   EST  01-OCT-2004
DEFINITION  43319.1 Common Scab-Challenged Tubers Solanum tuberosum cDNA clone
            43319 5', mRNA sequence.
ACCESSION   CV470446
VERSION     CV470446.1  GI:53695222
```

```
KEYWORDS   EST.
SOURCE     Solanum tuberosum (potato)
ORGANISM   Solanum tuberosum
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
           asterids; lamiids; Solanales; Solanaceae; Solanum.
REFERENCE  1  (bases 1 to 762)
AUTHORS    Flinn,B., Rothwell,C., Sardana,R., Griffiths,R., Lague,M., De
           Koeyer,D., Audy,P., Goyer,C., Li,X-Q., Wang-Pruski,G. and Regan,S.
TITLE      Generation of ESTs from common scab-challenged potato tubers
JOURNAL    Unpublished (2004)
COMMENT    Contact: Barry Flinn
           The Canadian Potato Genome Project - BioAtlantech
           921 College Hill Rd, Fredericton, ON, E3B 6Z9, CANADA
           Email: bflinn@bioatlantech.nb.ca
           Seq primer: T3.
FEATURES   Location/Qualifiers
    source  1..762
           /organism="Solanum tuberosum"
           /mol_type="mRNA"
           /cultivar="Shepody"
           /db_xref="taxon:4113"
           /clone="43319"
           /tissue_type="Tubers"
           /lab_host="XL10-Gold"
           /clone_lib="Common Scab-Challenged Tubers"
           /note="Vector: pBluescript II SK(+) XR; Site_1: EcoRI;
           Site_2: XhoI; supplier: Pathogen-challenge series. Tubers
           from pathogen-free Solanum tuberosum var. Shepody, clone
           1756, were inoculated with Streptomyces scabei EF-35, by
           applying 1 ml of a spore stock (OD650 = 0.7). Tubers were
           placed in trays of sand and watered from the bottom for
           the remainder of the experiment. Inoculated tubers were
           sampled at 2 days, 1 week, 2 weeks and 4 weeks
           post-infection. Harvested tubers were washed in distilled
           water, blotted dry, cut into small pieces and frozen in
           liquid nitrogen. RNA was isolated from each sample from
           collection, pooled and used for cDNA library
           construction."
ORIGIN
```

```
Query Match         9.5%;  Score 143;  DB 7;  Length 762;
Best Local Similarity  52.0%;  Pred. No. 1.5e-24;
Matches  345;  Conservative   0;  Mismatches  315;  Indels   3;  Gaps   1;

Qy   571 CAGTGGCTTTTTATCTGGAATGGGCCCATCTTTGTGGCCTTCGATTCCTGACCAAGGGGATT 630
            ||||    | |   ||||||
Db    98 CAGTTCTGTTCCAACGACAATGGAAAAAAAGTTCCTGACCATGCACGATCCAAAGGAATC 157

Qy   631 CAGATTATCGGCGATATGCCATCTTTGTGGCCTTCGATTCCTCGATTCGATGTCTGGGCCAAC 690
            ||||||   || |||      |||||| ||
Db   158 AGTATAATGGGAGACATGCCAATATATGTTGGATATCACAGTGCTGATGTTTGGGCCAAC 217

Qy   691 CCGCAGTACTTCTACCTCGAGGCCGATGGCAACCCACCGGTGTGGCGGCGGTTCCCGCGG 750
            || |     |||||| ||||||  |  |||| |||  ||||
Db   218 AAGAAACAATTTTGCTGAATAGGAAAAGGTTCCCTCTTATAGTTAGTGGTGGTCCTCCA 277

Qy   751 GACTACTTCTCCGAAACCGGCCAGCTCTGGCAATCCGCTCTATCCGCTGGGATGTGATG 810
            ||||   | ||  | |||       | ||||||
Db   278 GACGCCTTTAGTGAAACTGGTCAACTATGGGCCAGCCCTCTCTATGATTGGAAAGCCATG 337

Qy   811 GAAAGGGACAACTTTGCCTGGTGCATTGCCCGGACTGAGATTCGCTGCTTCAAGCAGTGCCAC 870
            ||||||       | |    ||||||  |||||||
Db   338 GAGAAGGATGGATTTCATGGTGGTATGCCAATCAACTGCAACCGGATCTTTTTTGAT 397

Qy   871 CTGGTGCCATGCAACCACTTCCGCGGGGTTGAAGCCTACTGAGGTTCCGTTTGGCCGG 930
            ||||||          |||||   ||
Db   398 GAATTTAGGATAGATCACTTTAGAGGATTTGCTGGTGATTTGGGCTTCCTTCTGAGGAA 457

Qy   931 CCCAATGCTGTGGAGGGGCTGGGTCAAAGCCCCAGGGAGAGAGAAGCTGTTGCTGCGGTG 990
            ||||      | ||||||  ||||   ||||||
Db   458 AAAATTGCAATTCGGGACGGTGGAAGGTGGACCTGAGAAAACCTTTGTTGGTTGATGCTATC 517
```

```
RESULT 6
AJ795827
LOCUS      AJ795827              733 bp  mRNA    linear    EST 08-DEC-2004
DEFINITION AJ795827 Antirrhinum majus whole plant Antirrhinum majus cDNA clone
           018_3_09_d02, mRNA sequence.
ACCESSION  AJ795827
VERSION    AJ795827.1  GI:51111155
KEYWORDS   EST.
SOURCE     Antirrhinum majus (snapdragon)
ORGANISM   Antirrhinum majus
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
           asterids; lamiids; Lamiales; Plantaginaceae; Antirrhineae;
           Antirrhinum.
REFERENCE  1  (bases 1 to 733)
AUTHORS    Bey,M., Stueber,K., Fellenberg,K., Schwarz-Sommer,Z., Sommer,H.,
           Saedler,H. and Zachgo,S.
TITLE      Characterization of Antirrhinum Petal Development and
           Identification of Target Genes of the Class B MADS Box Gene
           DEFICIENS
JOURNAL    Plant Cell 16 (12), 3197-3215 (2004)
PUBMED     15539471
COMMENT    Contact: Schwarz-Sommer Z
           Molekulare Pflanzengenetik
           MPI fuer Zuechtungsforschung
           Carl-von-Linne Weg 10, D-50829, Germany.
FEATURES   Location/Qualifiers
    source  1..733
           /organism="Antirrhinum majus"
           /mol_type="mRNA"
           /db_xref="taxon:4151"
           /clone="018_3_09_d02"
           /tissue_type="whole plant"
           /clone_lib="Antirrhinum majus whole plant"
ORIGIN
```

```
Query Match         9.3%;  Score 140.2;  DB 1;  Length 733;
Best Local Similarity  54.6%;  Pred. No. 7.4e-24;
Matches  280;  Conservative   0;  Mismatches  233;  Indels   0;  Gaps   0;

Qy   608 CCTATGCCGAATCCAAGGAGGGATTCAGATTATCGGCGAATATGCCCATCTTTGTGGCCTTCG 667
            ||||||   || ||||||  || |||||
Db   207 CTTATGCTCAGATGAGAGAGATCAGTATAATGCCTATATATGTGGGATATGC 266

Qy   668 ATTCCTCAGATGTCTGGGCCAACCCAGTACTTCTACCTCGAGGCCGATGGCCAACCCCA 727
            ||||||    |||||||||||
Db   267 ATAGTTCAGATGTGTGGGCTAATAAGAAAACATTTTTGCTTAATAGCAGTGCTGTTCCCC 326

Qy   728 CGGTGGTGGCGGCGGCCGTTCCGCGGGACTGTTTCGGTTTCGCCGGGTG 787
            |||   |   ||| ||||||   ||||  || ||  ||
Db   327 TTCTAGTTAGTGGTGGTCCCCTGATCCCTGATCCTCAGCGAGACTGGTCAGCTTTGGAACAGCC 386
```

Part (top of page, right column):

```
Qy   991 CGGGCCCAACTGAGCGATGCGCCCATCATTGCCGAAGACCTCGGGGTGATCACCCCGAG 1050
            |||||||   || |||||     ||||| |   |  | ||||||
Db   518 TTACAAGCTGTTGGGAAGATCAATATTATAGCAGAAGACTTGGGAGTAATTACCGAGGAC 577

Qy  1051 GTGGAGGCTTTGCGCGATGGCTTCGGGGTTCCCGGGTTCCCGATGAAGATTTTGCAGTTTGCTTTT 1110
            |||||||  |  |||| ||  ||||||||    ||
Db   578 GTTGTTCAGCTAAGAAAGTCCATTGAGCGACATGGCTGTGTACTCCAGTTTGCATT 637

Qy  1111 TCCGGTGAGGACGACAACGCCTTTGCCCCACCGACCCGGGATGGTTCCGCACCGGCAATGTGGTGGTG 1170
            |||||   |  || | |||
Db   638 GGCAGTGACGCGTGAAAACCCTCATTACCTCAATCATGAGCAGAACCAAGTA---GTG 694

Qy  1171 TACAGCGGAACCCACGACAACGACACACACCCTGGGATGGTTCCGACCGCGCCGGAGCC 1230
            |||||      | ||||| |||
Db   695 TATACTGGACAACATGACAATGATACGAGTGGTTGGTGGTGGGGATACTTTGCCCACAGGAA 754

Qy  1231 GAG 1233
            |||
Db   755 GAG 757
```

```
                                                                    method 4 described in Bonaldo et al., (1996) Genome
                                                                    Research 6: 791-806."

                                                          ORIGIN

                                                              Query Match          9.2%;    Score 138.2;   DB 3;   Length 538;
                                                              Best Local Similarity 56.4%;   Pred. No. 2.2e-23;
                                                              Matches    254;  Conservative    0;  Mismatches 196;   Indels    0;   Gaps    0;

Qy   788  CGCTCCTATCGCTGGGATGTGATGGAAAAGGACAACTTTGCCTGGTGCATTGCCCGGCATAA  847      Qy  468  GCCCTGGAACAGAGTGGAGCCCCGAGCTGCGCGACCGTGACCGGCTGCCCTGGCCAGGGC  527
          ||||||||| ||||||  ||||||||||| |||||| || |||||||||||||  ||||||                ||| | |||||| |||  | |  | | | ||||||||| ||||   ||| || | ||||||
Db   387  CTCTGTATGATTGGAAAGCCATGGAAATAGATGGATTTTCATGGTGGATAAAGCGTATAA  446      Db   82  GGCGGTGGTGGCAGTGGCCGGAGCCGTTGCGGCTGCGGCCAGAAGAGGCGCGTCAAGGAGTT  141

Qy   848  GGCAGTCGCTCAAGCAGTGCCACCTGGTGCGCATCGACCACTTCCGCGGGTTTGAAGCCT  907      Qy  528  CCGTGAGGAGCTGGCCGGAGAGGTGGCGCCCTTTACGAGTGGATTCAGTGGCTTTTTATCT  587
          ||| | ||||| | | |||| | ||| || ||||||||||||||||||||||| ||||||                ||| |  ||| ||||||| ||| ||| ||| |  | | |||| ||| | |||  || |||
Db   447  GACGTGCACAGAATCTATTCGATGAATTCCGGATCGACCACTTCAGAGGATTCGCTGGCT  506      Db  142  CCGGGAGACAAACAACAGGACGCAATTGACGAGTTCGTTGTGATCCAGTACTTCTTCGAGAA  201

Qy   908  ACTGGGAGGTTCCGTTTGGCCGCGGCCCAATGCTGTGGAGGGGCGCTGGGTCAAAGCCCCAG  967      Qy  588  GGAATGGGGCCAGACCAAAGCCTATGCGAATCCAAGGGATTCAGATTATCGGCGATAT  647
          |||| ||||||  |||||| |||| || ||| | ||| ||||| || || |||| ||                ||| ||| ||||  || ||| ||| || |  | | | |||| |||| ||| |||||||
Db   507  TTTGGGCTGTTCCTTCTGAAGCAAAAGTTGCGATGGTTGGGAAATGAAGGCAGGACCTG  566      Db  202  GCAGTGGAAGGCGATCCGGTCCTACGGGACGGCAAGGGCATCAAACTCATCGGCGACAT  261

Qy   968  GGGGAAAGCTGTTGTTTTGCTGCGCGTGCGGGCCCAACTGAGCGATGCGCCCATCATTGCCGAAG  1027     Qy  648  GCCCATCTTTGTGGGCCTTCAGTTCCTCAGATGTCTGGGCCAACCCGGGACTACTTCTACCT  707
          || ||| ||| ||| ||||| | || | |||||| ||||| | ||||| | | |||||||                ||||| || |  |||||| | |  | ||||||| |  || ||  | || ||||||| ||||
Db   567  GAGATTCTCATTTGATGCTATCTTTAGAGCTGTGTTGGAGATATCAATATCGTAGCAGAAG  626      Db  262  GCCCATCTACGTGGGCGGCCCACAGCCGCAGATGTGTGGGCCAACCGCCCACCTGTTCGAGCT  321

Qy  1028  ACCTCGGGGGTGATCACCCCGGAGGTGGAGCGCTTTGCGCGATGCGTTCCCGGTTCCCCCGGCA  1087     Qy  708  CGAGGCCGATGGCAACCCACGGTGGTGGCGGCGTTCGCGGGAGCAGGTCAGCGGAGTGGC  767
          ||||| |||| |  || |||||| |||||| | ||||| | |||| | | ||  ||||||                ||| || ||  || | |  | ||||| |  |||||| ||| ||| ||  ||| | || || ||
Db   627  ACTTGGGGGTCATTACTGAAGATGTAGTGCAGGCTTAGGACGTCATCGGAGCACCTGGAA  686      Db  322  GAACGAGGGCGGGCCTTGCCCGAGCAGCAGGTCAGCGAGTGTCGGCCGGACGCGGTCGCAAC  381

Qy  1088  TGAAGATTTTGCAGTTTGCTTTTTCCGGTGAGG  1120     Qy  768  CGGCCAGCTCTGGGCAATCCGCTCTATCGCCGTGGGGATGTGATGGAAAAGGGACAACTTTGC  827
          | |||||||| ||||| | | |||| ||||||                || ||| || ||||||||| |  ||| ||| ||  ||| ||||||| | |||||| ||||||
Db   687  TGGCCGTCCTCCAGTTTGGCTTCGGAAGTGATG  719      Db  382  GGGCCAGCTGTGGGGCAGCGGCTTTACAAGTGGCCGCGCCCACAAGAAGAGAGGGCTTCAA  441

                                                          Qy  828  CTGGTGCATTGCCCGCATAAGGCAGTCGCTCAAGCAGCAGTCGCTGCGCCATCGCCACCAGCGACCA  887
                                                                   || ||||| || || ||| || ||| | || || ||||| || ||  |  || ||||||||
                                                          Db  442  GTGGTGGACGGCGCGCATGGCCCGGACCCTGGAGCTGTACGACGAGTGNCNAATCGACCA  501

RESULT 7                                                   Qy  888  CTTCCCGGGTTTGAAGCCTACTGGAGGT  917
BI720945                                                            || |||||||| |||||| |||||||||
LOCUS       BI720945     538 bp   mRNA   linear   EST 19-SEP-2001    Db  502  CTTCCCGGTGGCTTCGCAGGCTACTGGGAGCGT  531
DEFINITION  1031053A09.y1 C. reinhardtii CC-1690, Stress II (normalized),
            Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.                RESULT 8
ACCESSION   BI720945                                                                   AY109140
VERSION     BI720945.1   GI:15696640                                          LOCUS       AY109140   1155 bp   mRNA   linear   HTC  25-FEB-2005
KEYWORDS    EST.                                                             DEFINITION  Zea mays PCO094517 mRNA sequence.
SOURCE      Chlamydomonas reinhardtii                                        ACCESSION   AY109140
  ORGANISM  Chlamydomonas reinhardtii                                        VERSION     AY109140.1   GI:21212571
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;            KEYWORDS    HTC.
            Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.            SOURCE      Zea mays
REFERENCE   1  (bases 1 to 538)                                                ORGANISM  Zea mays
  AUTHORS   Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C.,                   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Lefebvre,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,                   clade; Panicoideae; Andropogoneae; Zea.
            Unicellular System for Analyzing Gene Function and Regulation in  REFERENCE   1  (bases 1 to 1155)
            Vascular Plants. Project: 1031                                     AUTHORS   Gardiner,J., Schroeder,S., Polacco,M.L., Sanchez-Villeda,H.,
  JOURNAL   Unpublished (2001)                                                           Fang,Z., Morgante,M., Landewe,T., Fengler,K., Useche,F.,
COMMENT     Contact: Charles Hauser                                                      Hanafey,M., Tingey,S., Chou,H., Wing,R., Soderlund,C. and Coe,E.H.
            DCMB Box 91000                                                               Jr.
            Duke University                                                    TITLE     Anchoring 9,371 maize expressed sequence tagged unigenes to the
            Durham, NC 27708-1000                                                        bacterial artificial chromosome contig map by two-dimensional
            Tel: 919 613 8159                                                            overgo hybridization
            Fax: 919 613 8177                                                 JOURNAL   Plant Physiol. 134 (4), 1317-1326 (2004)
            Email: chauser@duke.edu.                                          PUBMED    15020742
FEATURES        Location/Qualifiers                                          REFERENCE   2  (bases 1 to 1155)
     source     1..538                                                         AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J,M., Whitsitt,M.S.,
                /organism="Chlamydomonas reinhardtii"                                    Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
                /mol_type="mRNA"                                               TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
                /strain="CC-1690 wild type mt+ 21gr"                                     Overgo Probes
                /db_xref="taxon:3055"                                          JOURNAL   Unpublished (2002)
                /clone_lib="C. reinhardtii CC-1690, Stress II               REFERENCE   3  (bases 1 to 1155)
                (normalized), Lambda Zap II"                                   AUTHORS   Coe,E.H.
                /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:      TITLE     Direct Submission
                XhoI; Stress condition II library, constructed by John        JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
                Davies and Jeffrey McDermott, combines cDNAs from CC-1690
                cells grown to mid-log phase in TAP (NH4+ - containing)
                and shifted to TAP - NO3- (24hrs); H2 production
                conditions (0, 12hr, 24hr) see Melis et al.,(2000) Plant
                Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
                sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
                PolyA mRNA was purified from each sample, pooled and cDNA
                synthesized. The cDNA was directionally cloned into lambda
                Zap II (Stratagene) in the EcoRI (5') and XhoI (3')
                sites. pBluescript II SK- plasmids were excised from the
                lambda ZAP clones by superinfection with ExAssist
                (Stratagene) phage. The library was normalized using
```

COMMENT　　　Missouri, Columbia, MO 65211, USA
　　　　　　　'If you are interested in getting corresponding physical clones,
　　　　　　　these are publicly available from ZmDB and may be found by BLAST
　　　　　　　searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
　　　　　　　www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
　　　　　　　maize cDNA sequences is either Virginia Walbot, Stanford or Pat
　　　　　　　Schnable, Iowa State, then clones may be requested from ZmDB:
　　　　　　　www.zmdb.iastate.edu.

FEATURES　　　Location/Qualifiers
　　　source　　　1..1155
　　　　　　　　　/organism="Zea mays"
　　　　　　　　　/mol_type="mRNA"
　　　　　　　　　/db_xref="MaizeGDB:635393"
　　　　　　　　　/db_xref="taxon:4577"
　　　　　　　　　/clone_lib="Maize Mapping Project/DuPont Cornsensus
　　　　　　　　　Library"
　　　　　　　　　/note="this sequence is part of a project of EST
　　　　　　　　　assemblies resulting from the application of public
　　　　　　　　　contigs to seed DuPont contigs; this resource was
　　　　　　　　　assembled by DuPont as part of a collaboration for the
　　　　　　　　　overgo addressing of BACs in conjunction with the Maize
　　　　　　　　　Mapping Project"

ORIGIN

Query Match　　　8.8%;　　　Score 132.4;　　DB 4;　　Length 1155;
Best Local Similarity　　52.2%;　　Pred. No. 7.1e-22;
Matches 318;　Conservative　　0;　Mismatches　288;　　Indels.　　3;　Gaps　　1;

Qy　624 GGGGATTCAGATTATCGGCGATATGCCCATCTTTGTGGCCTTCGATTCCTCAGATGTCTG 683
　　　　|||||||||||||||||||||||||| |||||| |||||||| |||||||||||||||||
Db　 53 GTGTATCAGCATCATGGGTGACATGCCTGTATATGTGGCTACCATAGCGGCATATGTTTG 112

Qy　684 GGCCAACCCGCAGTACTTCTACCTCGAGGCCGATGGCAACCCCACGGTGGTGGGCGGCGT 743
　　　　|||||||||||||||||||||||||| ||||| |||||||| |||||||||||||||||
Db　113 GGCGAACACGGAAAATCATTTTTGCTGGACAAAAACCGTTTCCCCACTTTCGTTAGTGGCGT 172

Qy　744 TCCGGGGACTACTTCTCCGAAACCGGCCAGCTCTGGGCAATCCGCTCTATCCGCTGGGA 803
　　　　||||||| || || ||||||||||| |||||| |||||| |||||||||||||||||
Db　173 TCCACCTGATGCATTTAGTGAAACGGGTCAACTATGGAACAGTCCATTGATTACGACTGGAA 232

Qy　804 TGTGATGGAAAGGACAACTTGCTGGTGCATTGCCCGCCATAAGGCAGTCGCTCAAGCA 863
　　　　||||||||||||||||| ||||||||||||| ||||||||||||||||||||||||
Db　233 AGCTATGGAAGCAGAGTTGTTTTCATGGTGGATAAAGAGGATTAAACGTGCCCTTGATT 292

Qy　864 GTGCCACCTGGTGCCCATCGCACTTCGCCGCGGGTTGAAGCCTACTGGGAGGTTCCGTT 923
　　　　|| |||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db　293 GTATGATGAATTCCGTATTGACCATTTCCGGGGCTTGCGGGTTTTTGGGCAGTCCCTTC 352

Qy　924 TGGCCGGCCCAATGCTGTGAGGGGCGCTGGGTCAAAGCCCCAGGGGAGAAGCTGTTTGC 983
　　　　|| ||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db　353 TGATGCAAAAGTAGCACTGGTTGGAAGCTCGGAGGGCTCGAGGCTCGAGGCTCTTTTTGA 412

Qy　984 TGCGGTGCGGGCCCAACTGAGCGATGCGCCCATCATTGCCGAAGACCTGGGGGGTGATCAC 1043
　　　　|| ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db　413 CACGGTCTTCAAAGCTGTTGGTAGAATAGATGATAATAGCAGAAGATCTCGGGGTAATTAC 472

Qy　1044 CCCGAGGTGGAGCTTGCGCGGATGGCTTCGGGTTCCCGGCATGAAGAGATTTTGCAGTT 1103
　　　　 ||| ||||||||||||||||||| ||||||||||||| |||||||||||||||||||
Db　 473 TGAAGATGTCGTTCAGCTAAGGAAATCCATTGGGCGCCCCTGGGATGGCAGTTCTGCCAGTT 532

Qy　1104 TGCTTTTTCCGGTGAGGACAACGCCTTTTGCCCACAACTACCCCGCACGGCAAATGT 1163
　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||
Db　 533 TGCTTTCGGAGGTGGTTCGACAACCTCATTGCCACAAACCATG---AAATGGATCA 589

Qy　1164 GGTGGTGTACAGCGGAACCCGGAACCCACAGACAACGACAACCCTGGGATGGTTCCGCACCGCGCC 1223
　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　 590 AGTTGTGTACACTGGAACATGATGATACAGTTCTTGGCTGGTGCAGCAAATTTACC 649

Qy　1224 GGAGGCCGA 1232
　　　　||||||||
Db　 650 AGAGGAAGA 658

──

LOCUS　　　　　CK767135
DEFINITION　　pam01-9ms3-e08 Pam01 Persea americana cDNA clone pam01-9ms3-e08 5',
　　　　　　　　mRNA sequence.
ACCESSION　　CK767135
VERSION　　　CK767135.1　GI:42721034
KEYWORDS　　　EST.
SOURCE　　　　Persea americana (avocado)
ORGANISM　　Persea americana
　　　　　　　Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
　　　　　　　Spermatophyta; Magnoliophyta; magnoliids; Laurales; Lauraceae;
　　　　　　　Persea.
REFERENCE　　1　(bases 1 to 678)
　AUTHORS　　dePamphilis,C., Carlson,J., Ma,H., Soltis,D., Soltis,P.,
　　　　　　　Oppenheimer,D., Frohlich,M., Doyle,J., Tanksley,S., Webb,M.,
　　　　　　　Leebens-Mack,J., Landherr,L., Ilut,D. and Wall,K.
　TITLE　　　Generation of ESTs from early flower buds of Persea americana
　JOURNAL　　Unpublished (2003)
COMMENT　　　Contact: Claude dePamphilis or James Leebens-Mack
　　　　　　　Mueller Laboratory
　　　　　　　Penn State University
　　　　　　　208 Mueller Laboratory, Department of Biology, ATTN Rm212, Penn
　　　　　　　State University, University Park, PA 16802, USA
　　　　　　　Tel: 814 863 6413
　　　　　　　Fax: 814 865 9131
　　　　　　　Email: cwd3@psu.edu or jhll0@psu.edu
　　　　　　　The sequence provided is trimmed of vector and low quality regions.
　　　　　　　Full sequence and original trace file are available from the Plant
　　　　　　　Genome Network website (http://pgn.cornell.edu)
　　　　　　　Plate: pam01-9ms3　row: e　column: 08
　　　　　　　Seq primer: M13F.

FEATURES　　　Location/Qualifiers
　　source　　　1..678
　　　　　　　　　/organism="Persea americana"
　　　　　　　　　/mol_type="mRNA"
　　　　　　　　　/db_xref="PGN:pam01-9ms3-e08"
　　　　　　　　　/db_xref="taxon:3435"
　　　　　　　　　/clone="pam01-9ms3-e08"
　　　　　　　　　/tissue_type="flower buds"
　　　　　　　　　/dev_stage="12-20 mm buds"
　　　　　　　　　/lab_host="SOLR"
　　　　　　　　　/clone_lib="Pam01"
　　　　　　　　　/note="Vector: pBluescript SK (+/-); Site_1: EcoRI;
　　　　　　　　　Site_2: XhoI; This is a directionally cloned,
　　　　　　　　　non-normalized library. This library has been generated by
　　　　　　　　　the Floral Genome Project (FGP). The Floral Genome Project
　　　　　　　　　is funded by NSF's Plant Genome Research Program
　　　　　　　　　(DBI-0115684). More information about the project can be
　　　　　　　　　obtained at http://fgp.bio.psu.edu"

ORIGIN

Query Match　　　8.5%;　　　Score 128.2;　　DB 7;　　Length 678;
Best Local Similarity　　52.9%;　　Pred. No. 7.2e-21;
Matches 322;　Conservative　　0;　Mismatches　283;　　Indels　4;　Gaps　　2;

Qy　632 AGATTATCGGCGATATGCCCATCTTTGTGCCCTTCGATTCCTCAGATGTCTGGGCCAACC 691
　　　　|||||||||||||||||||||||||||||| |||||||| |||||||||||||||
Db　 1 AGATCATGGGAGACATGCCTATTTATGTGGTTATCATAGTGCGGATGTCTGGGCAAACA 60

Qy　692 CGCAGTACTTCTACCTCGAGGCCGATGGCAACCCCACGGTGGTGGGCGGCGGTTCCGCGGG 751
　　　　|||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db　 61 AAAGGATATTTTTGCTGGACAAGAGTGGTTTCCCTACTCTTGTTAGCGGTGTTCCTCTG 120

Qy　752 ACTACTTCTCCGAAACCGGCCAGCTCTGGGCAATCCGCTCTATCGCTCG-GATGTGATG 810
　　　　|||||||||||||||||||||| |||||||||||||||||| ||||||
Db　121 ATGCCTTCAGTGAACATGGTCAACTATGGGACAGTCCCTGTATGACTGGAAAGGCCATG 180

Qy　811 GAAAGGGACAACTTTGCCTGGTGCATTGCCCGCCATAAGGCAGTCGCTCAAGCAGTCCAC 870
　　　　||||||||||||||||||||||||||||||||||||||||||||
Db　181 GAAGTTGATGGGTTTGCATGGTGGGTGAAGAGAATAAAGCGTGCACTTAATCTATATGAT 240

Qy　871 CTGGTGCGCATCGACCACTTCCGCGGGTTTGAAGCCTACTGGGAGGTTCCGTTTGGCCGG 930

```
                                                    Clontech's PCR Suppression Subtractive Hybridization was
                                                    used. Tester = plants sprayed with 50mM methyl jasmonate,
                                                    and tissue harvested 1,8, 24 and 48h after treatment;
                                                    driver = plants sprayed with water and tissue harvested
                                                    1,8, 24 and 48h. Cloned into Advantage PCR cloning vector
                                                    (Clontech)."

ORIGIN

        Query Match       8.4%;     Score 126.8;  DB 6;   Length 683;
        Best Local Similarity   54.1%;   Pred. No. 1.6e-20;
        Matches  331;   Conservative   0;  Mismatches 267;  Indels 14;  Gaps  3;

Qy   610 TATGCGAATCCAAGGGATTCAGATTATCGGCGATATGCCCATCTTGTGCGCCTTCGAT 669
         ||   ||||||||| | ||||  |  ||||  | ||||||  |||||| ||||| |||||
Db    71 TATGCACGGCGCAAGGAGTTGATATAAATGGAGATATGCCCATTTATGTAGGATATCAC 130

Qy   670 TCCTCAGATGTCTGGGCCAACCCCGAGTACTTCTACCTCGAGCGCATGGCAACCCCACG 729
         |||| ||||||| |||||| |||||||||| || || ||||| |||||| ||||  |||
Db   131 AGTGCAGACGTTGGGCAAATAAGAAACATTTCTTACTGACAAGAAGAAGGCTTTCCTCTT 190

Qy   730 GTGGTGGCGGGGCGTTCCGGCGGCTTCGGGCCAGCTCAGCTGGGCAATCCG 789
         || ||| ||||| || ||||| |||  |||||| |||||| |||||||| ||
Db   191 CTTGTTAGCGTGTGTTCCTCCTGACTTGTTCAGTGAAACTCGTCAACTGGT 250

Qy   790 CTCTATCGTCTGGGATGTGATGAAAGGACAACTTTGCCTCGTGCATTGCCGTCATAAGG 849
         || ||| | ||||| ||| ||||||| || ||| |||| || || || ||  ||||||
Db   251 CTTTATGACTGGAAAGCAATGGAGAGTGACCAATATTCTTGGTGGTTAATCGAATAAGA 310

Qy   850 CAGTCGCTCAAGCAGTGCCACCTGGTGCGCATCGACCACTTCCGCGGTTTGAAGCCTAC 909
         |  |||| | ||||| || || ||| || ||  |||| || |||| | ||| || ||
Db   311 CGCGCACAGGACTTGTATGACGAATGCAGGATTGATCACTTCAGAGGATTTGCAGGGTTT 370

Qy   910 TGGGAGGTTCCGTTGCGCCAATGCTGTGGAGGGGCCGCTGGGTCAAAGCCCCAGGG 969
         |||| ||| ||| |  | |||||| || |||||||||| || || |||| || |||
Db   371 TGGGCGGTCCCTTCTGAAGCGAAAGTTGCCATGGTTGGACATGG---AAGGACCTGGA 426

Qy   970 GAGAAGCTGTTTGCTGCGGTGGCGGGGCCCAACTGAGCGCATGAGCGCCATCATTGCCGAAGAC 1029
         |||||| || |||  | | | || |||||| ||||||| |||||| ||||| |||||| |||||
Db   427 AAGTCATTATTTGATGCGATGCCATTTCCAAAGCGTTGGGAAGATCAAAATCATAGCTGAGAT 486

Qy  1030 CTGGGGTGATCACCCCGAGTGGAGCTTTGCGCGATGGCTTCGGGTTCCCGGCATG 1089
         |||| || | ||| ||| || ||| |||||| || || || || |||||| ||||
Db   487 TTGGGAGTTATTACTAAAGATGTAGTTGAGCTGAGGAAATCTATCGGAGCACTGGAATG 546

Qy  1090 AAGATTTTGCAGTTGCTTTTTCCGGTGAGGAC---AACGCCTTTTTGCCCCACAACTA 1145
         |||| | |||| |||||| ||  |||| ||||    | |||| |||| |||||| |||
Db   547 GCCGTCCTCCAATTGCTTTGCTTTTGGAGGAGGCGGCCCGATAACCCACATTTACCTCACAATCA 606

Qy  1146 CCCCGCCACGGCAATGTGGTGGTTACAGCGAACCCACGACAACGAACACCCTGGG 1205
         || || |||| |||| ||| | || |||| |||||||| |||| || || ||||
Db   607 -----TGAAGTAAACCAAGTTGTATACTCTGGAAGTAACCACACACCACACTATTCGAAG 660

Qy  1206 ATGGTTCCGCAC 1217
         ||| |||| |||
Db   661 CTGGTGGGACAC 672
```

```
RESULT 10
CB074139
LOCUS       CB074139   683 bp   mRNA   linear   EST   01-JUL-2003
DEFINITION  EST01770 Methyl Jasmonate Treated Arabidopsis Forward-Subtracted
            Library Arabidopsis thaliana cDNA clone MJF-FB10 similar to
            4-alpha-glucanotransferase, mRNA sequence.
ACCESSION   CB074139
VERSION     CB074139.1  GI:32362520
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 683)
AUTHORS     Mahalingam,R., Gomez-Buitrago,A.M., Eckardt,N., Shah,N.,
            Guevara-Garcia,A., Day,P.M., Raina,R. and Fedoroff,N.V.
TITLE       Characterizing the stress/defense transcriptome of Arabidopsis
JOURNAL     Genome Biol. (2003) In press
COMMENT     Contact: Nina V. Fedoroff
            Biotechnology Institute
            Penn State University
            519 Wartik Lab, University Park, PA 16802, USA
            Tel: 8148635717
            Fax: 8148631357
            Email: nvf1@psu.edu
            Sequence does not include adaptor sequences (corresponding to
            Clontech PCR-Select Adaptors 1 and 2R) present on either side of
            insert.
            Insert Length: 683   Std Error: 0.00
            Seq primer: Clontech PCR-Select Nested Primer 2R
            High quality sequence start: 15
            High quality sequence stop: 675
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..683
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="MJF-FB10"
                     /tissue_type="Leaf"
                     /clone_lib="Methyl Jasmonate Treated Arabidopsis
                     Forward-Subtracted Library"
                     /note="Vector: Advantage PCR cloning vector (ClonTech);
```

```
Db   241 GAATTTCGGATAGACACCATTTTCGAGGACTTGCTGGCTTTTGGCGTGTTCCTTCTGAATCT 300
Qy   931 CCCAATGCTGTGGAGGGGCCGCTGGGTCAAAGCCCCAGGGGAGAAGCTGTTTGCTGCGGTG 990
Qy   301 AAGATTGCTATGTTCGGAAGATGGAAGCGTGGACCAAACAAAGCATTTTTTGATGCCGTT 360
Qy   991 CGGGCCCAACTGAGCCGATGCGCCCATCATTGCCGAAGACCTGGGGTGATCACCCCGAG 1050
Db   361 TTCAAAGCTGTTGGAAAGATCAATATATAATTGCTGAGGACTTGGAGTCATCACAGAGGAC 420
Qy  1051 GTGGAGGCTTTGCGCGATGGCTTCGGGTTCCCGGCATGGAAGATTTGCAGTTTGCTTTT 1110
Db   421 GTAATTCAGTTAAGAAAGCTATTGTGCACCAGGAATGGCAGTTCTCCAATTTGCTTTT 480
Qy  1111 TCCGGTGAGGACAACCCCGAGTACTTCTACCTCGAGCGCATGTGTGGTG 1170
Db   481 GGGAGTGATTCGACAATCCTCATTTACCTCATAATCATGAACATG---ATCAAGTCGTA 537
Qy  1171 TACAGCGGAACCCACGACAACGACACCCCCTCGGATGGTTCCGACACTGTCGTGGGAGGCC 1230
Db   538 TACACTGGAACACATGATGATAATGACACTGTCGTGGTGGTGGGAGAACTTGCCAAGAGGAA 597
Qy  1231 GAGCGGGCC 1239
Db   598 GAGAAGTCC 606
```

```
RESULT 11
CO119845
LOCUS       CO119845   646 bp   mRNA   linear   EST   16-JUN-2004
DEFINITION  GR_Eb022P17.r GR_Eb Gossypium raimondii cDNA clone GR_Eb022P17
            3', mRNA sequence.
ACCESSION   CO119845
VERSION     CO119845.1  GI:48818532
KEYWORDS    EST.
SOURCE      Gossypium raimondii
ORGANISM    Gossypium raimondii
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1 (bases 1 to 646)
AUTHORS     Kim,H., Yu,Y., Kudrna,D., Hatfield,J., Stum,D., Mueller,C.,
            Udall,J.A., Rapp,R.A., Wendel,J.F., Rao,K., Soderlund,C. and
```

TITLE   Wing,R.A.
            Global assembly of Cotton ESTs
JOURNAL   Unpublished (2004)
COMMENT   Contact: Rod A. Wing
            Arizona Genomics Institute
            The University of Arizona
            Forbes Building Room 303, Tucson, AZ, 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: rwing@genome.arizona.edu
            Plate: 022 row: P column: 17.
FEATURES   Location/Qualifiers
     source   1..646
                     /organism="Gossypium raimondii"
                     /mol_type="mRNA"
                     /db_xref="taxon:29730"
                     /clone="GR_Eb022P17"
                     /tissue_type="floral"
                     /dev_stage="-3 to +3 DPA"
                     /lab_host="DH10B"
                     /clone_lib="GR_Eb"
                     /note="Vector: pCMV.SPORT-6.1; Site_1: NotI; Site_2:
                     EcoRV; Library made by Invitrogen with RNA supplied by
                     Wendle lab. Directional cloned into NotI-EV. Colonies
                     plated/picked by AGI. More glycerol clones held in -80."

ORIGIN

Query Match   8.3%; Score 125.4; DB 7; Length 646;
Best Local Similarity 51.9%; Pred. No. 3.6e-20;
Matches   282; Conservative   0; Mismatches 261; Indels   0; Gaps   0;

Qy   571 CAGTGGCTTTTTATCTGGAATGGGGCCAGACCAAGGCCTATGCCGAATCCAAGGGGATT 630
         |||||| || || | | |||| || | | | | || | | ||| |
Db   66 CAATTCCGTTCCAAAGGCAGTGGAAGAAAATTCGCGACTATGCACAGACAAAGGGAGTC 125

Qy   631 CAGATTATCGGCGATATGCCCATCTTTGTGGCCTTCGATTCGTCAGATGTCTGGGCCAAC 690
         | |||||| || | ||||| || || || || | || | |||||| ||| |||
Db   126 AGCATAATGGGAGACATGCCTATTTATGTAGGTTATCACAGTGCTGATGTTTGGGCAAAT 185

Qy   691 CCGCAGTACTTCTACCTCGAGGCCGATGGCAACCCCACGGTGGTGCGGGGTTCCGCGG 750
         | |||| ||| || | || | ||| | |||| | ||| || || |||| | |
Db   186 AAGAAACATTTTTGCTGAATAGGCATGGTTATCCTATGCTAGTCAGTGGTGTTCCTCCT 245

Qy   751 GACTACTTCTCCGAAACCGGCCAGCTCTGGGCAATCCGCTCTATCCGGCTGGGATGTGATG 810
         || | ||| ||| ||| |||| || || | || || || ||| || ||| |||| |||
Db   246 GATGCCTTCAGTGATACTGGTCAGCTTCGTGGGGAGTTCCTCTATATGATTGGAAAGCCATG 305

Qy   811 GAAAGGGACAACTTTGCCTGGTGCATTGCCCGCATAAGGCAGTCGCTCAAGCAGTGCCAC 870
         || | | | ||| || | | |||| | ||| | | || | ||| || | |
Db   306 GAGAAAGATGGATTTTTATGTGGATAAGTCGCATTAGACGTGCACAGACTCTATATGAT 365

Qy   871 CTGGTGCGCATCGACCACTTCCGCGGGTTTGAAGCCTACTGGGAGGTTCCGTTTGCCGG 930
         ||| | | | ||||| || | |||||| ||| || |||| | ||| | ||| ||
Db   366 GAATTCAGGATAGATCACTTTAGAGGATTTCTGTGGGCTTTTCGGCGTGCTGGTCCGTT 425

Qy   931 CCCAATGCTGTGGAGGGGGCGCTGGGTCAAAGCCCCAGGGGAGAAGCTGTTTGCTGCGGTG 990
         | | ||| || | | || | || || || | || || ||| ||| | | |
Db   426 AAAGTTGCAATGGTTGGACGATGGAAGGTTGTCCTCAAAAAATCTTTGTTGTTGACGCGCATC 485

Qy   991 CGGGCCCAACTGAGCCGATGCCCCATCATTGCCGAGACCTGGGGGTGATCACCCCCGAG 1050
         | ||| |||| | ||| || || | || ||| ||||| || | | |||
Db   486 TCCAAAGCAGTTGGAAAGATTAATATCATAGCTGAAGATCTGGGAGTTTATAACTGAGGAT 545

Qy 1051 GTGGGAGGCTTTGCGCGATGGCTTCGGGTTCCCGGTTTGCAGTTTGCTTT 1110
         || | || ||| || || || || | | ||| | || |||| || | ||
Db   546 GTAGTTCAGCTCAGAGAAATATATTGGAGCACCTGGAATGGCTGTGTTGTTTG 605

Qy 1111 TCC 1113
         |||
Db   606 TCC 608

RESULT 12

BE515486/c
LOCUS   BE515486   605 bp   mRNA   linear   EST 08-AUG-2000
DEFINITION WHE0613_B06_C11ZA Wheat ABA-treated embryo cDNA library Triticum
            aestivum cDNA clone WHE0613_B06_C11, mRNA sequence.
ACCESSION   BE515486
VERSION   BE515486.1 GI:9739432
KEYWORDS   EST.
SOURCE   Triticum aestivum (bread wheat)
  ORGANISM Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1 (bases 1 to 605)
  AUTHORS   Anderson,O.D., Chao,S., Han,P.S., Hsia,C.C., Johnson,R.R, Kang,Y.,
            Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L., Tong,J.C.,
            Verhey,S.D. and Walker-Simmons,M.K.
  TITLE   The structure and function of the expressed portion of the wheat
            genomes - ABA-treated embryo library
  JOURNAL   Unpublished (2000)
COMMENT   Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Clontech Matchmaker 3' AD primer.
FEATURES   Location/Qualifiers
     source   1..605
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Brevor (soft, white, winter, common wheat)"
                     /db_xref="taxon:4565"
                     /clone="WHE0613_B06_C11"
                     /tissue_type="Seed embryo"
                     /dev_stage="Mature dormant seeds"
                     /lab_host="E. coli DH12S"
                     /clone_lib="Wheat ABA-treated embryo cDNA library"
                     /note="Vector: pGAD10; Site_1: EcoRI; Site_2: XhoI;
                     Embryos were cut from mature, dormant seeds and imbibed in
                     25 microM ABA (abscisic acid) in 5 mM Mes buffer, pH 5.7,
                     for 12 hr at 22 C. The tissue, total RNA, and poly(A)
                     RNA were prepared by Steven Verhey in M.K.
                     Walker-Simmons's lab (USDA-ARS, Washington State Univ.,
                     Pullman, Washington 99164-6420. A cDNA library was made
                     by Clontech using a combination of random and oligo dT
                     primers. Library was plated and archived by Russell
                     Johnson (Colby College, ME/Walker-Simmons' lab). Plasmid
                     DNA preparations and DNA sequencing were performed in the
                     OD Anderson lab (all other authors)."

ORIGIN

Query Match   8.0%; Score 120; DB 2; Length 605;
Best Local Similarity 52.1%; Pred. No. 7.9e-19;
Matches   267; Conservative   0; Mismatches 245; Indels   0; Gaps   0;

Qy   415 TGGCTGGAAGACTATGCGCTCTTTATGGCCCTCAAGACACCGGTTTGACGGCAAGCCCTGG 474
         ||||| ||||| | | | | ||| | | | | | | ||| ||| || |||| |||
Db   553 TGGCTTGAAGATGCTGACTTTCGCGCTATCGACATAGCACAGTCAGTGTCCTGG 494

Qy   475 AACGAGTGGAGCCCCGAGCTGCGCGACCGTGAACCGGCTGCCCTGGCCCAGGGCCCGTGAG 534
         || ||||||| ||| || | | || | || | | || | || ||| || || ||| |||
Db   493 TCTGAGTGGCCTGAACCGGTGAAAGATCGCCATCCTGGAGCCTTGGAGCCTTGAAAGATATATGAA 434

Qy   535 GAGCTGCCGAGGAGGTGGGCCCTTTACGAGTGGATTCAGTGGCTTTTTATCTGGAATGG 594
         || | || || |||||||| |||||| | |||| | | | ||| ||| |||||
Db   433 AACCAGAAGGACTTTATAGAAAATTTATGGCACAACAGTTCTTATTTGAAAAGCAATGG 374

Qy   595 GGCCAGGACCAAGGCCTATGCCGAATCCAAGGGGATTCAGATTATCGGCGATATGCCCATC 654
         | ||| ||| ||| |||||| || | || |||| ||| | || | | ||||| |||
Db   373 AAACGGGTTCGTTCACATGCACAGAAGCTTGCATCACAGACATAATGGGTGACATGCCAATA 314

RESULT 12

```
Qy  751 GACTACTTCTCCGAAACCGGCCAGCTCTGGGGCAATCCGCTCTATCGCTGGGATGTGATG 810
        ||| |||       ||||  ||||| ||     ||  |   ||||| ||||||||| |||
Db  220 GATGCATTTAGTAGAACGGTCAACTATGAACAGTCCATTGTACGACTGGAAAGCTATG 279

Qy  811 GAAAGGGACAACTTTGCTGGTGCATGCCCGCATAAGGCAGTCGCTCAAGCAGTGCCAC 870
        ||| |  |  | ||| | ||| | ||||        ||||| |  ||| | |||  |
Db  280 GAAGCAGATTGTTTTCATGGTGGATAAAGAGGATTAAACGTGCCCTTGATTTGTATGAT 339

Qy  871 CTGGTGCGCATCGACCACTTCCGCGCGGTTTGAAGCCTACTGGGAGGTTTGGCGCGG 930
        |  | | ||| |||||| ||||| | | |||||    |    | | ||| || ||
Db  340 GAATTCCGTATTGACCATTTCCGGCGGCTTGCTGGGGTTTTTGGGCAATCCCTTCTGATGCA 399

Qy  931 CCCAATGTCTGTGGAGGGGCGCTGGGTCAAAGCCCAGGGGAGAAGCAGTGTTTGCTGCCGTG 990
        |||||  || | || | | | ||| | |||| ||| ||||  || || |  | | | ||
Db  400 AAAGTAGCACTGGTTGGAAGCTGGAGGGCTGGACCAAGGAATAGCTTTTTTTGACACGCTC 459

Qy  991 CGGGCCCAACTGAGCGAATGCGCCCATCATTGCCGAAGACCTGGGGTGATCACCCCCGAG 1050
        | || | ||| ||| | |  |   | || || | |||| | ||||  | |   |||| |||
Db  460 TTCAAAGCTGTTGGTAGAATAGATATAATAGCAGAAGATCTCGGGGGTAATTACTGAAGAT 519

Qy 1051 G 1051

Db  520 G 520
```

RESULT 14
CF437795
LOCUS   CF437795   765 bp   mRNA   linear   EST 04-SEP-2003
DEFINITION   EST674140 normalized cDNA library of onion Allium cepa cDNA clone
             ACAEB58, mRNA sequence.
ACCESSION   CF437795
VERSION   CF437795.1 GI:34460485
KEYWORDS   EST.
SOURCE   Allium cepa (onion)
ORGANISM   Allium cepa
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Alliaceae;
             Allium.
REFERENCE   1 (bases 1 to 765)
AUTHORS   Havey,M.J., Cheung,F., Van Aken,S., Utterback,T. and Town,C.D.
TITLE   Expressed Sequence Tags from a normalized library of mixed onion
             tissues (Allium cepa)
JOURNAL   Unpublished (2003)
COMMENT   Contact: Havey MJ
             Department of Horticulture
             USDA-ARS and University of Wisconsin
             1575 Linden Drive, Madison, WI 53706, USA
             Tel: 608-262-1830
             Fax: 608-262-4743
             Email: mjhavey@facstaff.wisc.edu
             TIGR sequence name ACAEB58TR. For more information:
             http://haveylab.hort.wisc.edu
             Seq primer: CAG GAA ACA GCT ATG ACC.
FEATURES   Location/Qualifiers
     source   1..765
                     /organism="Allium cepa"
                     /mol_type="mRNA"
                     /cultivar="Red Creole (bulbs), unknown(callus), Ebano &
                     Texas Legend(roots)"
                     /db_xref="taxon:4679"
                     /clone="ACAEB58"
                     /tissue_type="Callus, roots, and young bulbs"
                     /clone_lib="normalized cDNA library of onion"
                     /note="Vector: pCMVSport6.1-ccdb (Invitrogen); Site 1:
                     EcoRV (5'); Site 2: NotI (3'); Equal molar amounts of mRNA
                     from callus, roots, and young bulbs were combined to
                     synthesize the library. Normalization to enrich for
                     low-copy transcripts was performed by proprietary
                     techniques of Invitrogen."
ORIGIN

Query Match   7.8%;   Score 117.4; DB 6; Length 765;
Best Local Similarity 50.3%;   Pred. No. 3.6e-18;

```
Qy  655 TTTGTGCCTTCGAATTCCTCAGATGTCTGGGCCAACCCGCAGTACTTCTACCTCGAGGCC 714
        | ||| | |||| ||  || |||||| |||||  |    ||||  | ||    |  ||
Db  313 TATGTTGGCTACCACAGCGCAGATGTTTGGGCAAACAGGAAATCATTTTGCTGGACAAT 254

Qy  715 GATGGCAACCCCACGGTGGTGCGGGCGTTCCGCGGGACTACTTCTCCGAAACCGGCCAG 774
        ||||   |||||| | ||| | || ||||| | ||||||| ||       ||||| ||
Db  253 AATGGTTTCCCCACCATTTGTTAGTGGTGTTCCACCTGATGCGGTTCAGTAAAACTGTTCAG 194

Qy  775 CTCTGGGGCAATCCGCTCTATCGCTGGGATGTGATGGAAAGGGACAACTTTGCTGGTGC 834
        || ||| |||  |   ||||| ||||||| | |  |||| | ||  |||    |||
Db  193 TTATGGAACAGCCCCATTGTATGACTGAAATCTATGGAAGCAGATGGCTTTGCATGGTGG 134

Qy  835 ATTGCCCGCATAAGGCAGTCGCTCAAGCAGTGCCACCTGGTGCGCATCGACCACTTCCGC 894
        | | |   ||| |  | | |   || || ||    | || | | || |||| ||| ||||
Db  133 GTAAAGAGGATTAAAACGTGCCCTTGATTTGTATGAGTTCCGTATTGACCCATTTCCGC 74

Qy  895 GGGTTTGAAGCCTACTGGGAGGTTCCGTTTGG 926
        ||| ||| |||  |  |||  |||| |  ||
Db  73  GGGCTTGCTGCTGGTTTTTGGGCAGTTCCTTCGG 42
```

RESULT 13
CF013402
LOCUS   CF013402   548 bp   mRNA   linear   EST 17-JUL-2003
DEFINITION   QBK5e12.xg QBK Zea mays cDNA clone QBK5e12, mRNA sequence.
ACCESSION   CF013402
VERSION   CF013402.1 GI:32908589
KEYWORDS   EST.
SOURCE   Zea mays
ORGANISM   Zea mays
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
             clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 548)
AUTHORS   Genoplante.
TITLE   Genoplante, a major partnership french program in plant genomics
JOURNAL   Unpublished (2003)
COMMENT   Contact: Genoplante
             Genoplante
             93, rue Henri Rochefort 91025 EVRY CEDEX France
             Tel: 33 1 69 47 54 00
             Fax: 33 1 69 47 54 10
             This sequence has been generated in the framework of the french
             plant genomics programme 'Genoplante' (http://www.genoplante.com
             and http://genoplante-info.infobiogen.fr).
FEATURES   Location/Qualifiers
     source   1..548
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /cultivar="F2"
                     /db_xref="taxon:4577"
                     /clone="QBK5e12"
                     /tissue_type="embryo"
                     /clone_lib="QBK"
ORIGIN

Query Match   7.8%;   Score 117.8; DB 6; Length 548;
Best Local Similarity 52.8%;   Pred. No. 2.7e-18;
Matches 254;   Conservative 0;   Mismatches 227;   Indels 0;   Gaps 0;

```
Qy  571 CAGTGGCCTTTTTATCTGGAATGGGCCAGACCAAGGCCTATGCCGAATCCAAGGGGATT 630
        |||| |  ||| |   ||| | |||||  | ||| || ||| ||| |||  || | ||
Db  40  CAGTTCTTATTTCAAAGGCAATGGCAGCGGATTCGTAAATATGCACARAAAGCTGGGTATC 99

Qy  631 CAGATTATCGGCGATATGCCCATCTTTGTGGCCTTCGATTCCCAGATGTCTGGGCCAAC 690
        ||||| || ||| |||||||| | | || ||  | |||  |   || |  ||||| ||
Db  100 AGCATCATGGGTGACATGCCTATATATGTTGGCTACCATAGCCGCAGATGTTTGGGCGAAC 159

Qy  691 CCGCAGTACTTCTACCTCGAGGCCGATGCAACCCACGGTGGTGCGGGCGGTTCCGCGG 750
        | ||| | || || | | || |   ||   |||| | || | |  | |||| ||| |
Db  160 AGGAAATCATTTTGCTGGACAAAAAACGGTTTCCCCACTTTCGTTAGTGGCGTTCCACCT 219
```

Matches   289;   Conservative   0;   Mismatches   286;   Indels   0;   Gaps   0;

```
Qy    415  TGGCTGGAAGACTATGCGCTCTTATGGCCCTCAAGACCCGGTTTGACGGCAAGCCCTGG  474
                |||||||||||||  ||  |    |   |||       | | |    ||||||||||
Db    104  TGGCTTGAAGATGCTGCACTTTTTTCAGCTATTGACCATAGCATTGATACATTTTCCTGG  163

Qy    475  AACGAGTGGAGCCCCGAGCTGCGCGACCGTGAACCCGGCTGCCCTGGCCCAGGGCCCTGAG  534
                | |||||          |    |||  | ||| |     | | ||     |||
Db    164  ACTGAGTGGCGCTGAACCTCTTAAAAATCGTCATCTTGATGCTGTTTGGAAGAGCATATCGA  223

Qy    535  GAGCTGGCCGGAGGAGTGGGCCCTTTACGAGTGGATTCAGTGGCTTTTTATCTGAAATGG  594
                |  ||     |    |  |  | |  | |   | | |   ||||    |||||
Db    224  ATCCACGAGACTATATAGACATTTTTGTTGCCCAACAATTTTTGTTTCAAAAACAATGG  283

Qy    595  GGCCAGCCAAGGCCTATGCCGAATCCAAGGGGATTCAGATTATCGCGGATATGCCCATC  654
                |||||  | |||   | |   | || |||||     ||  | ||||||||||||
Db    284  GTAAACATTCACAAATCGCGCGACACAAACTAGGAATCAAAATTATGGTGATATGCCGATT  343

Qy    655  TTTGTGGCCTTCGATTCCTCAGATGTCTGGGCCAACCCGCAGTACTTCTACCTCGAGGCC  714
                ||| |  |||||| |   | |  ||   || ||||||||||  |  | ||||||
Db    344  TATGTTGGTCATCAAAGTGCAGATGTGTGGGCAAATAAGAAATCCTTAGTGCTGAACAGG  403

Qy    715  GATGCAACCCCACGGTGGTGGCGGGCGTTCCGGCGGACTACTTCTCCGAAACCGGCCAG  774
                ||||||||| |||| | |  |||| || |||  | | |  ||||||| ||||||| ||
Db    404  ATTGGTTATCCAGTTCGTTGTCAGTGCGTTCCGCCTGATTGTTTTAGTGAAACTGGTCAA  463

Qy    775  CTCTGGGCAATCCGCTCTATCGCTCTGGGATGCTGATGGAAAGGGACAACTTTGCCTGGTGC  834
                ||||||||||||| |  | | ||   |    | |||||||| |   |  |||||  | ||
Db    464  TTATGGGGAAGTCCTCTATATGATTGGAGGGCAATGGAGGGCAATGATATGCTTGGTGG  523

Qy    835  ATTGCCCGCATAAGGCAGTCGCCTCAAGCAGTGCCACCTGGTGCGCATCGACCACTTCCGC  894
                |||||  |||| | || |||||  | | ||||| | ||||   |    | || ||||||
Db    524  ATAAAAGAATAAAACGTGCTGTTCCTTGATCTGTATGATGAGTTCGCATAGATCACTTCCGT  583

Qy    895  GGGTTTGAAGCCTACTGGGAGGTTCCGTTTGCCCGGCCCAATGCTGTGGAGGGGCGCTGG  954
                ||| |||||||     || | |   ||||   ||||  || |||| || |||  ||||
Db    584  GGACTAGCTCGGATACTGGGCCTTCGTGCCTTCCGAAGCAAAGTCGCAATGTTTGGAAGATGG  643

Qy    955  GTCAAAGCCCCAGGGGAGAAGCTGTTTGCTGCGGT  989
                || |||| || |||||||||||  | ||||| ||||
Db    644  AAGGCTGGACCAGGGGAAAACTTTCTTTGATGCTGT  678
```

ORIGIN

Query Match   7.8%;   Score 116.6;   DB 7;   Length 769;
Best Local Similarity 50.1%;   Pred. No. 5.8e-18;
Matches 358;   Conservative   0;   Mismatches 339;   Indels   18;   Gaps   2;

```
Qy    716  ATGGCAACCCCACGGTGGTGGCGGGCGTTCCGGCGGACTACTTCTCCGAAACCGGCCAGC  775
                ||||||||     ||   ||||   || |  |  || |  ||||||  |||||| |||
Db     17  AAGGTTTCCCTCTTATAGTTAGTGGTGTTCCTCCAGACGCCTTTAGTGAAACTGGTCAAC   76

Qy    776  TCTGGGCAATCCGCTCTATCGCTGGGATGTGATGGAAAGGGACAACTTTGCCTGGTGCA  835
                |||||||||||| | ||  | |    |  | |||||| |  |  |   ||||| | |||
Db     77  TATGGGGCAGCCCTCCTATGATGGAAAGCCATGGAAGGATGGATTTTCATGGTGGG  136

Qy    836  TTGCCCGCATAAGGCAGTCGCTCAGCAGTGCCACCTGGTGCGCATCGACCACTTCCGCG  895
                ||||   ||| | || ||||||   |||  |||    |  ||| | |||| ||| |||
Db    137  TATGCCGAATTCAACGTGCAACGACGATCTTTTTGATGAATTTAGGATGATCACTTTAGAG  196

Qy    896  GGTTTGAAGCCTACTGGGAGGTTCCGTTTGGCCGGCCCAATGCTGTGGAGGGGCGCTGGG  955
                |||| ||||||   |||||| ||||| || || || | |||||||||| || | ||||
Db    197  GATTTGCTGGATTTTGGGCTTGTTCCTTCTGAGGAAAAAATTGCAATTCTGGACGGTGGA  256

Qy    956  TCAAAGCCCCAGGGGAGAAGCGTGTTTGCTGCGGGCCCAACTGCAGCCGATGCGCCCA  1015
                |||||  ||| | || ||||| ||||  || | | | || |||| ||||||||||||
Db    257  AGGTGGGACCTGGAAAACCTTGTTTGATGCGTATCTTCTACAAGCTGTGTGGAAGATCAATA  316

Qy   1016  TCATTGCCGAAGACCTGGGGGTGATCACCCCCGAGTGGAGGCTTTGCGCGATGGCTTCG  1075
                |||||||||||||   || |||| |  |  |||||||||||||||    || |  |||
Db    317  TTATAGCAGAAGACTTGGGAGTAATTACCGAGGACGGTTGTTCAGCTGAAGAAAGTCCATTG  376

Qy   1076  GGTTCCCCGCATGAAGAATTTGCACTTCAGTTTGCATTTGGCAGTGACGGCTGGAAACCCTCATT  1135
                ||||||| ||| ||||| |  | | ||  |||||  |||||||   | || ||||||
Db    377  AGGGACCTGGAATGGCGCACGGCAATGTGGCAATGTGGTGTACAGGGGAACCCACGACAACGA  436

Qy   1136  CCCACAACTACCCCCGCGCACGGCAATGTGGCAATGTGGTGTACAGGCGGAACCCACGAACAACAGACA  1195
                |||||  || || ||     ||  |       |||||||||||||||
Db    437  TACCTCACAATCATGAGCA---GAACCAAGTAGTGTATACTGGAACACATGACAATGATA  493

Qy   1196  CCACCTGGATGGTTCCGCACCGGCCGAGCGGCCTTCATGCGGGCCTACC  1255
                || |||||||  ||  |||||||||||  |  || |  ||||||||||
Db    494  CGATCCGAGGTTGGTGGGATACTTTGCCACAGGAAGAAATCCAATGTACTAAAGTATT  553

Qy   1256  TGGCCCGCTATGGCATCCGCATCCGTCGTTGTTGTTGCTGCGGCCGCTTTGATCGAGC  1315
                ||||||  || || | |||||||| |||||||||||||||  |||||||||||
Db    554  TATC--------AAATATTGAGGAAGAGGAAATATCATGGGGCTTGATCGAAG  598

Qy   1316  TGGCCTTCAAAAGCCCGGCCAAGCTGGCTATTGTGCCTTTGCGGGACGTGCTGGGGCTGG  1375
                ||||   || | |    || |  | |  ||||||  |||| |  | |   || |||||
Db    599  GTGCAGTTGCTTCTTCTGTAGCCCGTATTGCAATTATACCGATGCAGGATGTGTTCTTGGGCTTG  658

Qy   1376  GCCCCGAGGCCCGCATGAACTTCCCCGACGGCTGGGGGACAACTGGGCGTGGCG  1430
                |||||||||||   ||||| | |||||||| ||||||  | ||||  |||||||
Db    659  GGAGTGATTCAGAATGAACATTCCAGCAACTCAGTTTGGAAACTGGGAGTGGAG  713
```

Search completed: January 14, 2006, 14:44:25
Job time : 5560 secs

RESULT 15
CV302316
LOCUS   CV302316   769 bp   mRNA   linear   EST 23-SEP-2004
DEFINITION 67345.1 Swollen Stolon Solanum tuberosum cDNA clone 67345 5', mRNA
            sequence.
ACCESSION   CV302316
VERSION   CV302316.1 GI:52619649
KEYWORDS   EST.
SOURCE   Solanum tuberosum (potato)
ORGANISM   Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum.
REFERENCE   1 (bases 1 to 769)
AUTHORS   Flinn,B., Rothwell,C., Sardana,R., Griffiths,R., Lague,M.,
            DeKoeyer,D., Audy,P., Goyer,C., Li,X.-Q., Wang-Pruski,G. and
            Regan,S.
TITLE   Generation of ESTs from swollen stolon tissues of potato
JOURNAL   Unpublished (2004)
COMMENT   Contact: Barry Flinn
            The Canadian Potato Genome Project - BioAtlantech
            921 College Hill Rd, Fredericton, ON, E3B 6Z9, CANADA
            Email: bflinn@bioatlantech.nb.ca
            Seq primer: T3.
FEATURES   Location/Qualifiers
     source   1..769
                     /organism="Solanum tuberosum"
                     /mol_type="mRNA"
                     /cultivar="Shepody"
                     /db_xref="taxon:4113"

```
            GenCore version 5.1.6
            Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2006, 10:41:10 ; Search time 300 Seconds
                                      (without alignments)
                                      8905.584 Million cell updates/sec

Title:          US-10-618-976-1
Perfect score:  1503
Sequence:       1 atgcaactccaacgcgctt........ccgaggccagccagccgcgct 1503

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters:   2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents NA:*
             1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
             2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
             3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
             4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
             5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
             6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
             7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
             8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
             9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1503	100.0	1503	3	US-09-687-360-1	Sequence 1, Appli
	2	289	19.2	1518	3	US-09-583-110-1567	Sequence 1567, App
	3	287.4	19.1	5910	3	US-08-961-527-165	Sequence 165, App
	4	265.6	17.7	1377	3	US-09-107-433-336	Sequence 336, App
	5	113.2	7.5	1485	3	US-09-902-540-2554	Sequence 2554, Ap
c	6	113.2	7.5	15377	3	US-09-902-540-1116	Sequence 1116, Ap
	7	64.2	4.3	1557	3	US-09-719-978-1	Sequence 1, Appli
	8	47.2	3.1	2930	3	US-09-902-540-9600	Sequence 9600, Ap
	9	47.2	3.1	12194	3	US-09-902-540-1091	Sequence 1091, Ap
c	10	46.4	3.1	1683	2	US-08-838-543-1	Sequence 1, Appli
c	11	46.4	3.1	1877	3	US-09-780-173A-10	Sequence 10, Appl
	12	46.2	3.1	11220	3	US-09-105-537-32	Sequence 32, Appl
	13	46.2	3.1	36778	3	US-09-105-537-5	Sequence 5, Appli
	14	46.2	3.1	38506	3	US-09-320-878-19	Sequence 19, App
	15	46.2	3.1	38506	3	US-09-141-908-1	Sequence 1, Appli
	16	46.2	3.1	38506	3	US-09-657-440-19	Sequence 19, Appl
	17	46.2	3.1	38506	3	US-09-793-708-19	Sequence 19, Appl
	18	46	3.1	552	3	US-09-003-287-3	Sequence 3, Appli
	19	44	2.9	1776	3	US-09-902-540-7046	Sequence 7046, Ap
c	20	43.2	2.9	4532	3	US-09-902-540-626	Sequence 626, App
	21	43.2	2.9	2967	3	US-09-185-501B-12	Sequence 12, Appl
	22	42.8	2.8	1287	3	US-09-252-991A-5766	Sequence 5766, Ap
	23	42.8	2.8	1329	3	US-09-252-991A-5804	Sequence 5804, Ap
c	24	42.8	2.8	1857	3	US-09-252-991A-5847	Sequence 5847, Ap
	25	42	2.8	666	3	US-09-902-540-9322	Sequence 9322, Ap
	26	42	2.8	9839	3	US-09-902-540-996	Sequence 996, Appl
	27	41.8	2.8	5811	3	US-10-152-886-2	Sequence 2, Appli
	28	41.4	2.8	699	3	US-09-252-991A-9676	Sequence 9676, Ap
	29	41.4	2.8	1062	3	US-09-902-540-3602	Sequence 3602, Ap
	30	41.4	2.8	1371	3	US-09-252-991A-9717	Sequence 9717, Ap
	31	41.4	2.8	20113	3	US-09-902-540-1173	Sequence 1173, Ap
	32	40.8	2.7	1208	2	US-08-403-852D-4	Sequence 4, Appli
	33	40.8	2.7	1208	3	US-08-510-646B-4	Sequence 4, Appli
	34	40.8	2.7	1208	3	US-09-231-818-4	Sequence 4, Appli
	35	40.8	2.7	1208	3	US-09-635-359B-4	Sequence 4, Appli
	36	40.8	2.7	5392	3	US-08-403-852D-1	Sequence 1, Appli
	37	40.8	2.7	5392	3	US-08-510-646B-1	Sequence 1, Appli
	38	40.8	2.7	5392	3	US-09-231-818-1	Sequence 1, Appli
	39	40.8	2.7	5392	3	US-09-635-359B-1	Sequence 1, Appli
c	40	40.6	2.7	360	3	US-09-252-991A-6224	Sequence 6224, Ap
	41	40.6	2.7	633	3	US-09-252-991A-6234	Sequence 6234, Ap
	42	40.6	2.7	987	3	US-09-758-759-70	Sequence 70, Appl
	43	40.6	2.7	1029	3	US-09-252-991A-6399	Sequence 6399, Ap
	44	40.6	2.7	1050	3	US-09-252-991A-10218	Sequence 10218, A
c	45	40.6	2.7	1122	3	US-09-252-991A-9836	Sequence 9836, Ap

ALIGNMENTS

```
RESULT 1
US-09-687-360-1
; Sequence 1, Application US/09687360
; Patent No. 6617143
; GENERAL INFORMATION:
; APPLICANT: Fukuyama, Shiro
; TITLE OF INVENTION: Polypeptides Having Glucanotransferase
; TITLE OF INVENTION: Activity And Nucleic Acids Encoding Same
; FILE REFERENCE: 6012.200-US
; CURRENT APPLICATION NUMBER: US/09/687,360
; CURRENT FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: PA 1999 01501
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: PA 1999 01641
; PRIOR FILING DATE: 1999-11-15
; PRIOR APPLICATION NUMBER: 60/160,903
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: 60/166,539
; PRIOR FILING DATE: 1999-11-18
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1503
; TYPE: DNA
; ORGANISM: Thermus rubens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1503)
US-09-687-360-1

Query Match             100.0%;  Score 1503;  DB 3;  Length 1503;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 1503;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy     1 ATGCAACTCCAACGCGCTTTGGGCCGCGAGGCCGAGCGGGTTTTCCACCCACCAGTTTCCGGGTCGCTGG  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 ATGCAACTCCAACGCGCTTTGGGCCGCGAGGCCGAGCGGGTTTTCCACCCACCAGTTTCCGGGTCGCTGG  60

Qy    61 GGGATTGGGGCTCTGGGCCCGCGAGGCCGAGCGGGTTTTGGACTGGCTGGCCGATGCGGGGA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 GGGATTGGGGCTCTGGGCCCGCGAGGCCGAGCGGGTTTTGGACTGGCTGGCCGATGCGGGGA 120

Qy   121 GCCCCTGGTGGCAGGTCTTACCGCTGGGCCCTACCAGTTACGGGGACTCCGCGTACCAG 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GCCCCTGGTGGCAGGTCTTACCGCTGGGCCCTACCAGTTACGGCGACTCCGCGTACCAG 180

Qy   181 TCCTTCTCCGGCTTTTGCCGGTAACCCGTATTGGTTGACCCGAGATGCTGATTGAAAAA 240
```

RESULT 2
US-09-583-110-1567
; Sequence 1567, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 1567
; LENGTH: 1518
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-1567

Query Match   19.2%;   Score 289; DB 3; Length 1518;
Best Local Similarity 51.7%; Pred. No. 4.5e-64;
Matches 764; Conservative 0; Mismatches 699; Indels 18; Gaps 4;

Qy   11 AACGCGCTTTCGGAATTTTGCTCACCCCACCAGTTTCCGGTTCGCTCGGGGGATTGGGG 70
           |||||| |||| || ||||| || ||||||| ||||| |||||||| |||||||||
Db   8 AACGTCAAAGTGGTGTTGTTGATGCACATCTCTTCTTCCAGGAGCTTACGGAATCGGAT 67

Qy   71 CTCTGGGCGCCGAGGCCGGTTTTGGACTGGCCGATGCGGGAGCCCGCTCGT 130
           ||||| |||||||| | || |||| |||||| |||| | | ||| | ||||
Db   68 CATTTGGTCAAAGTGCTTACGACTTCGTTGATTTCGTTCGTCGTCACAACGTTACT 127

Qy   131 GGCAGGTCTTACCGCTGGGCCCTACCAGTTACGGCGACTACCGCCGTACCAGTCCTTCTCGG 190
           ||||| ||||| ||||| || ||||| ||||| ||||| || |||| ||||| |||||||
Db   128 GGCAAATCCTTCCATTAGGACGCAACTAGTTACGGGGATTCTCCTTACCAATCTTTCTCAG 187

Qy   191 CTTTTGCCGGTAACCCGTATTTCGGTGGTGCCATTCATTTATCGATTTATCGATTCGTGGTTG 250
           |||| ||||||| |||||| ||||| |||||| | |||||||||| ||||| |||||||||
Db   188 CCTTCGCAGGAAACACTCATTTATCGATTTATCGATTTAGAATCTTGGTGGAGCAAGGTTGTTGG 247

Qy   251 AACAAAGCGAAGCGCCCCCGTATCCGACCACCACC 310
           ||| ||| ||||||||||||||| ||||||||||||
Db   248 AAGCAAGTGACCTTGAAGGAGTTGGTAGCGATGCGTCTGAAGTTGACTATGCTA 307

Qy   311 AGACCCGCTGGCCCGTTGCGCGCGGGCTTCGGCCC 370
           ||||||||| | |||||||||||||| ||||| ||
Db   308 AAATCTACTATGCACGTGCTCCTCTTTTAGAAAAAGCGGTGAAACGTTTCTTTGAAGTCG 367

Qy   371 AGGATAAGACCCGACTGGAAGCCTTATCGAGCGGAAGACTATG 430
           |||||| |||||||| |||||||||| ||||||||||||||||

[DNA sequence alignment region — left column, Db/Qy pairs]

Db   181 TCCTTCTCGGTTTTGCCGGTAACCCGTATTTGGTGACCCCGAGATGCTGATTGAAAAA 240
Qy   241 GGCTGGCTGGAACAAAGCGAAGCGCCCCCGTATCCGACCACCAGCCGTGGATTATGGC 300
Db   241 GGCTGGCTGGAACAAAGCGAAGCGCCCCCGCCCGTATCCGACCACCAGCCGTGGATTATGGC 300
Qy   301 TGGCTTTACCAGACCCGCTGGCCCCTGTTGCGCGGGCTTTCGCGGGGTTTCGGGCAAGG 360
Db   301 TGGCTTTACCAGACCCGCTGGCCCCTGTTGCGCGGGCTTTCGCGGGGTTTCGGGCAAGG 360
Qy   361 GCTTCGGCCCAGGATAAGACCCGACTGGAAGCCTTTATCGAGGCGCGTTCGGCTG 420
Db   361 GCTTCGGCCCAGGATAAGACCCGACTGGAAGCCTTTATCGAGGCGCGTTCGGCTG 420
Qy   421 GAAGACTATGCGGCTCTTTATGCCCTTCAAGACCCGTTTGACGGCAAGCCCTGGAACGAG 480
Db   421 GAAGACTATGCGGCTCTTTATGCCCTTCAAGACCCGTTTGACGGCAAGCCCTGGAACGAG 480
Qy   481 TGGAGCCCGAGCGTGGCCCTGAACCGGTGCCCTGGCCAGGGCCCGTGAGGAGCTG 540
Db   481 TGGAGCCCGAGCGTGGCCCTGAACCGGTGCCCTGGCCAGGGCCCGTGAGGAGCTG 540
Qy   541 GCCGAGGAGTGGCCCCTTACGAGTGGCTTTTTATCTGGAATGGGGCCAG 600
Db   541 GCCGAGGAGGTGGCCCCTTACGAGTGGCTTTTTATCTGGAATGGGGCCAG 600
Qy   601 ACCAAGGCCTATGCCCAAGGGATTCAGATTATCGGCGATATATGCCATCTTTGTG 660
Db   601 ACCAAGGCCTATGCCCAAGGGATTCAGATTATCGGCGATATATGCCATCTTTGTG 660
Qy   661 GCCTTCGATTCCTCAGATGTCTGGGCCAACCCGAGTACTTCTACCTCGAGGCCGATGGC 720
Db   661 GCCTTCGATTCCTCAGATGTCTGGGCCAACCCGAGTACTTCTACCTCGAGGCCGATGGC 720
Qy   721 AACCCCACGGTGGTGCGGGGGTTCGCGGGGACTACTTCTCCGAAACCGGCCAGCTCTGG 780
Db   721 AACCCCACGGTGGTGCGGGGGTTCGCGGGGACTACTTCTCCGAAACCGGCCAGCTCTGG 780
Qy   781 GGCAATCCGCTCTATCCGCTGGATGTGATGAAAAGGACAACTTTGCCTGTGCAATTGCC 840
Db   781 GGCAATCCGCTCTATCCGCTGGATGTGATGAAAAGGACAACTTTGCCTGTGCAATTGCC 840
Qy   841 CGCATAAGGCAGTGCGCTCAAGACAGTGCCACCTGGTGCGCATCGACCACTTCCGCGGGTTT 900
Db   841 CGCATAAGGCAGTGCGCTCAAGACAGTGCCACCTGGTGCGCATCGACCACTTCCGCGGGTTT 900
Qy   901 GAAGCCTACTGGGAGGTTCCGTTGCGCGCGGCCAATGCGTGGAGGGGCGCGTGGGGTCAAA 960
Db   901 GAAGCCTACTGGGAGGTTCCGTTGCGCGCGGCCAATGCGTGGAGGGGCGCGTGGGGTCAAA 960
Qy   961 GCCCCAGGGGAGGAGAGCTGTTTGCTGCGGGTGCGGCCGATGAGCGGCCCATCATT 1020
Db   961 GCCCCAGGGGAGGAGAGCTGTTTGCTGCGGGTGCGGCCGATGAGCGGCCCATCATT 1020
Qy 1021 GCCGAAGACCTGGGGTGATCACCCCCGAGGTGGAGGCTTTGCGGGAGGCTTCGGGTTC 1080
Db 1021 GCCGAAGACCTGGGGTGATCACCCCCGAGGTGGAGGCTTTGCGGGAGGCTTCGGGTTC 1080
Qy 1081 CCCGGCATGAAGATTTTGCAGTTTGCTTTTTTCCGGTACAGCGGCAACCGGCCTTTTGCCCCAC 1140
Db 1081 CCCGGCATGAAGATTTTGCAGTTTGCTTTTTTCCGGTACAGCGGCAACACGCCTTTTGCCCCAC 1140
Qy 1141 AACTACCCCGCCACGGCAATGTGGTGGTGTACAGCGGACCGGCCTTACCTGGCC 1200
Db 1141 AACTACCCCGCCACGGCAATGTGGTGGTGTACAGCGGAACCGGCCTTACCTGGCC 1200
Qy 1201 CTGGGATGGTTCCGCACCGGCCGAGGCCTTCATGCGGGCCTTACCTGGCC 1260
Db 1201 CTGGGATGGTTCCGCACCGGCCGAGGCCTTCATGCGGGCCTTACCTGGCC 1260
Qy 1261 CGCTATGGCATCCGTTGTTGTCGGAATACGAGGTCGCGGGGCCTTTGATCGAGCTGGCC 1320

[DNA sequence alignment region — right column]

Db 1261 CGCTATGGCATCCGTTGTTGTCGGAATACAGAGGTCGCGGGGCCTTTGATCGAGCTGGCC 1320
Qy 1321 TTCAAAAGGCCCGGCCAAGCTGGCTATTGTGCCTTTGCCTTTGCCGAGGACGTGTCGGGCTGGGCGGGCCCC 1380
Db 1321 TTCAAAAGGCCCGGCCAAGCTGGCTATTGTGCCTTTGCCTTTGCCGAGGACAACTGGGCGTGCGGGGCCCC 1380
Qy 1381 GAGGCCCGCATGAACTTCCCGGACGGCTTGGGGCACAACTGGGCGTGCGCTACGCCGAA 1440
Db 1381 GAGGCCCGCATGAACTTCCCGGACGGCTGGGGGACTGGGCCCTGCGGGAGCACTGGGCGTGCGCTACGCCGAA 1440
Qy 1441 GGCGACCTCGAGCCCGGGGACTGCGGGCCCTGCGCCGAGGGCCAGCAGCGC 1500
Db 1441 GGCGACCTCGAGCCCGGGGACTGCGGGCCCTGCGCCGAGGGCCAGCAGCGC 1500
Qy 1501 GCT 1503
           |||
Db 1501 GCT 1503

RESULT 3
US-08-961-527-165
; Sequence 165, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Human Genome Sciences, Inc.
;   STREET: 9410 Key West Avenue
;   CITY: Rockville
;   STATE: Maryland
;   COUNTRY: USA
;   ZIP: 20850
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;   COMPUTER: HP Vectra 486/33
;   OPERATING SYSTEM: MSDOS version 6.2
;   SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/961,527
;   FILING DATE:
;   CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Brookes, A. Anders
;   REGISTRATION NUMBER: 36,373
;   REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (301) 309-8504
;   TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 165:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 5910 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
US-08-961-527-165

Query Match   19.1%; Score 287.4; DB 3; Length 5910;
Best Local Similarity 51.7%; Pred. No. 1.9e-63;
Matches 763; Conservative 0; Mismatches 696; Indels 18; Gaps 4;

```
Qy        11 AACGCGCTTTTGGAATTTGCTCCACCCACCAGTTTCCGGTCGCTGGGGATTGGGG     70
              ||||||| |||| ||| |||| |  |  |||  |  |||||||| ||  |||| |||||||
Db       108 AACGTCAAAGTGGTGTGTTGATGCACATCTCTTCTTCCAGGAGCTTACGGAGGAATCGGAT    167

Qy        71 CTCTGGGCCGGAGCCGGAGCGGTTTTGGACTGGCTGGCCGGCCGATGCGGGAGCCCGCTGGT    130
             ||||  |||  | || || |  |  | || || ||| ||  ||| ||||| ||  ||  |
Db       168 CATTTGGTCAAAGTGCTTACGACTCGTGTGATTTCTTCGGTGTTACGTACACAAACGTTACT    227

Qy       131 GGCAGGTGCTTACCGCTGGGGCCTGGGGCCGACTCCGGTACGGCGGCCTCCTTCTCGG    190
              ||| || || | ||  || ||||  ||  ||  || | || || ||  |  |||||
Db       228 GGCAAATCCTTCCATTAGGAGCAACTAGTTACGGGGATTCTTCCTTACCAATCTTTCTCAG    287

Qy       191 CTTTTCGCGTAACCGTATTGGTGACCCGGAGATGCTGATTGAAAAAGGCTGGCTGG    250
             |||  ||||  |  ||  ||  | ||  || ||| | ||| |  |  ||  |  ||||
Db       288 CCTTCCGCAGGAAACACTCATTTTATCGATTTAGATATCTTCCGGCAGTTGGTTGTTGG    347

Qy       251 AACAAAGCGAAGCGCCCCGCCGTATCCCCAGCACCCCAGGGATCGATTATGGCTCGATTATCGACCGTGA    310
             || || || ||| |||  ||| ||  ||   || || || ||||| |||  |||
Db       348 AAGCAAGTGACCTTGAAGGACGGTCTCGAAGGCTTGGGGGGTTTCGGGCAAGGGTTCGGCCC    407

Qy       311 AGACCCGCTGGCCCCGTGTTGCGGCCCTGGGGGTTTCGGGCAAGGGTTCGGCCC    370
              || ||| |   | |||  |||  ||  ||  |  | ||
Db       408 AAATCTACTATGCACGTCGTCCTCTTTTAGAAAAAGCGGCGCTTCTCGCTGAAGACTATG    467

Qy       371 AGGATAAGACCCGACTGGAAGCCTTTATCGAGGCCGGAGACGTATG    430
Db       1433 TGACACCAGCTGTCGAGGAAGGTTTGTCTTGACTTGAC    1469
```

Db   1533 TGACACCAGCTGTCGAGGAAGAGTTGCTTGCTTGACTTGAC 1569

RESULT 4
US-09-107-433-336
; Sequence 336, Application US/09107433
; Patent No. 6800744
GENERAL INFORMATION:
    APPLICANT: Lynn A Doucette-Stamm and David Bush
    TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
                       SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
                       THERAPEUTICS
    NUMBER OF SEQUENCES: 5206
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: GENOME THERAPEUTICS CORPORATION
        STREET: 100 Beaver Street
        CITY: Waltham
        STATE: Massachusetts
        COUNTRY: USA
        ZIP: 02354
    COMPUTER READABLE FORM:
        MEDIUM TYPE: CD/ROM ISO9660
        COMPUTER: <Unknown>
        OPERATING SYSTEM: <Unknown>
        SOFTWARE: <Unknown>
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/09/107,433
        FILING DATE: 30-Jun-1998
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: 60/ 085131
        FILING DATE: May 12, 1998
        APPLICATION NUMBER: 60/051553
        FILING DATE: July 2, 1997
    ATTORNEY/AGENT INFORMATION:
        NAME: Ariniello, Pamela Deneke
        REGISTRATION NUMBER: 40,489
        REFERENCE/DOCKET NUMBER: GTC-011
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: (781)893-5007
        TELEFAX: (781) 893-8277
INFORMATION FOR SEQ ID NO: 336:
    SEQUENCE CHARACTERISTICS:
        LENGTH: 1377 base pairs
        TYPE: nucleic acid
        STRANDEDNESS: double
        TOPOLOGY: circular
    MOLECULE TYPE: DNA (genomic)
    HYPOTHETICAL: NO
    ANTI-SENSE: NO
    ORIGINAL SOURCE:
        ORGANISM: Streptococcus pneumoniae
    FEATURE:
        NAME/KEY: misc feature
        LOCATION: (B) LOCATION 1...1377
    SEQUENCE DESCRIPTION: SEQ ID NO: 336:
US-09-107-433-336

Query Match   17.7%; Score 265.6; DB 3; Length 1377;
Best Local Similarity   52.1%; Pred. No. 4.2e-58;
Matches 698; Conservative   0; Mismatches 624; Indels 18; Gaps   4;

Qy 328 TTGCGGGGGCTTTCGGGGTTTCGGGCAAGGGCTTCGGCCCAGGATAAGACCCGACTG 387
         ||| | ||||||||| ||||||||| || | |||||||| || ||
Db 184 CGTCCCTTTTAGAAAAAGCGGTGAAACGTTTCTTTGAAGTCGGAGATGTTAAAGATTTT 243

Qy 388 GAAGCCTTTATCGAGGCCGAGCGCTTCTGGCTGGAAGACTATGCGTCTCTTTATGGCCCTC 447
         || | |||| || | ||||| ||||||| || | ||| ||| ||
Db 244 GAGAAATTTGCTCAAGACAACCAATCATGGCTTGAGCTCTTTGCTGAGTATATGGCTATC 303

Qy 448 AAGACCCGGTTTGACGGCCAAGCCCTTGAACGAGTGAGC---CCCGAGCTGCGCGACCGT 504
         || | ||||||||| | |||||||| ||| | ||| |||| |||| ||
Db 304 AAAGAGCATTTTGACAATCTTGCTTGGACTGAATGGCCAGATGCAGATGCTCGTGCTCGT 363

Qy 505 GAACCGGCTGCCCTGGCCCAGGGCCCGTGAGGAGCTGGCCCCTTTACGAG 564
         |||||| ||||| ||| ||| |||| |||| ||||| |||| || |
Db 364 AAAGCTTCAGCACTTGAAAGCTATGCTGAGCAATTGGCAGAGCAAGTTGGTTTACCACCGT 423

Qy 565 TGGATTCAGTGGCTTTTTTATCTGGAATGGGGCCAGACCAAGGCCTATGCCGAATCCAAG 624
         || | || ||| | ||||| | ||| |||||| |||| || || || ||||| |||
Db 424 GTGACTCAATACTTCTTCTTCAACAATGGTTGAAATTGAAAGCTTACGCTAACGACAAC 483

Qy 625 GGGATTCAGATTATCGGCGATATGGCCCATCTTGTGCGCCTTCAGATGTCTGG 684
         ||| || |||| | || ||| |||||| || || ||| || |||| |||
Db 484 CACATCGAAATCGTTGGGGACATGGCAATTGGCTTGTTTGTTCAAGTTGATGTGTGG 543

Qy 685 GCCAACCCGCAGTACTTCTACCTCGAGGCCGATGGCAACCCCACGGTGGTGCGGGCGTT 744
         |||||| ||||| || ||| || |||| || |||| |||| || | |||| ||
Db 544 GCAAATCCACCATCTCTTCAAAAACAGATGTCAATGGTAAGGCTACTTGTATCGCAGATGC 603

Qy 745 CCGCGGGACTACTTCTCCGAAACCGGCAGCTCTGGGGCAATCCGCTCTATCGTCTGGGAT 804
         || || |||||| | |||||| |||| ||| ||| |||||| | | | ||| ||
Db 604 CCACCAGATGAGTTTTCTGTAACTGGTCAGCTTTGGGGTAATCCAATCTATGACTGGGAA 663

Qy 805 GTGATGGAAAGGGACAACTTTGCTCGGTCGTCGTGCATTGCCCGCGATAAGCGTGAAAATC 864
         || || |||| |||| |||| | | | ||| || | ||| | || || | |
Db 664 GCAATGGACAAAGACGGCTACAAAATGGTTGAATGGCTACCGGCCCCATCGAACTTCAAAATC 723

Qy 865 TGCCACCTGGTGCGCATCGACACTT 924
         |||||| || || || ||||
Db 724 TACGATAATCGTTCGTATCGACCAC 783

Qy 925 GGCCGGCCCAATGCTGTGGAGGGGCTGAGTAGTCCGAGGAGAAGCTGT 984
Db 724 TACGATAATCGTTCGTATCGACCACTTCCGCTGGTGCTGGCTGCTTACTACTGGCTTT 783

Qy 985 GCGGTGCGGGCCCAACTGCCGATGAGCGATGCCGAGCTGGTGGGTGATCACC 1044
Db 784 GGTTCCGATACAGCAGCACCTGGTAGTGAGTGGTACAGCTTTTTGCA 843

Qy 1045 GCCGTTAAGGAAGAAGAACTTGGTGAGCTAAAACATCATCCGAGAAGCACCTTGAGCAGTTT 1104
Db 844 GCCGTTGCGGTGCGGGCCCAACTGCCGATGAGCGATGCGGAGCTGGTGGGTGATCACC 903

Qy 1105 GCTTTTCCGGTGAGGACGCAACGCCTTTTGCCCCACAACTACCCCACACT---TGGCACCTGCTAACTCA 1164
Db 904 GATGAAGTGATGATCGAAATTGCCGTTCGTTCTTGTTGTACCGTAATGCAATGTG 963

Qy 1165 GTGGTGTACAGCGGACGTGCTGGGGCTGGGGATGGTTCCGCACCGCCCG 1224
Db 964 GCCTTCAACCCGGAAGATGCGAAAGCATTGATAGCCCCACCT---TGGCACCTGCTAACTCA 1020

Qy 1225 GAGGCGAGCGGCCTTCATGCGGGGCTTACCTGCCCG 1284
Db 1021 GTTATGTACACAGGAACACACGATAACAATACGGTTCTTGTGGTTGTACCGGTAATGAGATT 1080

Qy 1285 GAATACGAGGTCGCGGGGTCGGCCTTCAAAAAGCCCGGGCCAAGCTGGCT 1344
Db 1081 GATGATGCG-----------ACTCGTGAGTAGTACATGGCTGCCTTTGATCGTCGTGCTTGTTCACCGA 1131

Qy 1345 ATTGTGCCTTTGCAGGACGTGCTGGGGCTGGGGCTGCCGTATGGAATTGGCA 1404
Db 1132 GAAACAGTGGT---ACACGCTATGCTTCGTTACATGGCTTTTTCATCGTTTATGGCA 1188

Qy 1405 CGGCTGGGGCGGCAACTGGGCGTGGTGCGCTACGGCGAGAGGCGACCTTCGAGCCCGGTCTGGCC 1464
Db 1189 ATTGCAACTATGCAAGATTTACTAGAATTGGATGAGGCAGCTTCGTATGAAGTTCCCCATCT 1248

RESULT 5
US-09-902-540-2554
; Sequence 2554, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 2554
; LENGTH: 1485
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-2554

Query Match   7.5%; Score 113.2; DB 3; Length 1485;
Best Local Similarity 53.8%; Pred. No. 4.2e-19;
Matches 269; Conservative 0; Mismatches 213; Indels 18; Gaps 1;

Db 1249 ACCCTTGGTGGAAACTGGTCTTGGCGTATGACTGAGATCAATTGACACCAGCTGTGTCGAG 1308
         ||||| || ||||| |||| || ||||| || ||| |||| | || ||||| ||||
Qy 399 CGAGGCCGAGCGCTTCTGGCTGGAAGACTATGCGTCTCTTTATGGCCCTCAAGACCCGGTT 458

Db 1465 GCGGGACTGCGGGGCCCTGGC 1484
         |||||| || |||| ||||
Db 378 CGACGCCCAGGGCGAATGGCTGGAGAGCTACCGGCTACCGCCTCTTCACCGCGACCAGGA 437

Qy 459 TGACGGCAAGCCCGCCGGCGCCGGAGCGAGTGAGCCCCGAGCTGCGCGACCGGCTGCCCT 518
Db 438 AGAGCGCCGCCCGCCGGCGCCCGTGGGAGTGGCCGGAGGAGCGGGTGCGCGCACCCGAGCCCT 497

Qy 519 GGCCAGGCGCCCGCGAGGAGCTGGCCCAGGAGGAGCGGGCCTTTACGAGTGGATTCAGTGGCT 578
Db 498 GGCCCAGAAGTCCCGGAGTCGGCCGGAGCGCGGTGAGCGGGAGCTACCAGGTCCCAGTGGGT 557

Qy 579 TTTTATCTGGAATGGGGCCAGACCAAGGCCTATGCGCAAGGGGATTCAGATTAT 638
Db 558 GGCCGAGCAGCAGTGGAACGAGGTGCGGACGCGCAGGGCCCGCGCAAGGACGTCTCCTGTG 617

Qy 639 CGGCGGATATGCCCATCTTGTGCGTCGGTTCGATTCCTCAGATGTCTGGGCCAACCCGCAGTA 698
Db 618 CGGCGACGAGCCGTTCATCATCCGGTGGGCGCGCGACGTGTGGGCGCCACCGGACACTT 677

Qy 699 CTTCTACCTCGAGGCCGATGGCAACCCCACGGTGGTGCGGGCGTTCCGCGGGACTACTT 758
Db 678 CCTGCGCGGCGACGCGC-------------CGCTCGGGCGCGTGCCGCTCCGGATGGACTT 719

Qy 759 CTCCGAAACCGGCAGCTCTGGGGCAATCCGCTCTATCGTCTGGGATGTGATGGAAAGGGA 818
Db 720 CTCCGCCACAGGGCGCCAGGACTGGGCGCCTGCTGCCTCCACTTGACCTCAAGCAGTAGGAGGA 779

Qy 819 CAACTTTGCTCGGTGCTATTGCCCGCATAAGGCCCGGCGATTACACGACAACCGGTAAAGAATAC 878
Db 780 CGACTTCGCGGCTGGGCGGAGGCGCGGAGAAGGCGGCCAGCTACTACGACTTGCCGCCG 839

Qy 879 CATCGACCACTTCCGCTGGGT 898
Db 840 CGTGGACCACCGGGTGGGGGT 859

RESULT 6
US-09-902-540-1116/c

; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 1
; LENGTH: 1557
; TYPE: DNA
; ORGANISM: Chlamydomonas reinhardtii
US-09-719-978-1

Query Match   4.3%;   Score 64.2;   DB 3;   Length 1557;
Best Local Similarity   53.6%;   Pred. No. 1.5e-06;
Matches 157;   Conservative   0;   Mismatches 133;   Indels   3;   Gaps   1;

Qy   49 CCGGGTCGCTGGGGATTGGGGCTCTGGCCCGCGAGGCCGAGCGGTTTTGGACTGGCTG 108
          |||||| ||| || ||||||| || || |||| |||||||| ||||| ||||||||
Db   1 CCGGGTGCGTACGGTATTGGCCGATATTGGTGATGGTGAGGCCAAGCGCCTTCGTCGACTGGCTC 60

Qy   109 GCCGATGCGGGAGCCCGCTGGTGGCAGGTCTTACCGTCGGGCCCTACCAGTTACGG---C 165
          ||||| ||||||||||| || |||| || ||||||| ||||| |||||||
Db   61 GCCGACCACGGCATGCAGTCGTGGCAGCTGCTGCTCCCCCGGACCCCATGTAC 120

Qy   166 GACTCGCCGTACCAGTCCTTCTCGCTTTTCGCGGTAACCCGTATTTGGTTGACCCGGAG 225
          ||||| |||||||||| |||||| || |||| ||||||||||
Db   121 TACTCCCCTACTCGGGACACGGAGGCCAACTGCGGCAACCCCTTGTTGGTGAGCATTGAG 180

Qy   226 ATGCTGATTGAAAAGGTGGCTGGACAAAGCGAAGCGCCCCGCGCGTATCCGACCCAG 285
          ||||||| || ||||| |||||| || |||| || |||||||
Db   181 GAGCTCATCAAGGACGGACTGCTGGAGTTCTCGGAGACGCCCGGTGCCCCATCGCC 240

Qy   286 CGCGTGGATTATGGCTGGCTTTACCAGACCCGCTGGCCCTGTTGCGGCGGGC 338
          |||| || ||||| ||| ||||||||| || ||||||
Db   241 GACGTTGACTATCCCGCTGTGGCCGCCGCCAAGCTGCCCTGCTCAAGCGCGC 293

RESULT 8
US-09-902-540-9600
; Sequence 9600, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 9600
; LENGTH: 2930
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-9600

Query Match   3.1%;   Score 47.2;   DB 3;   Length 2930;
Best Local Similarity   51.5%;   Pred. No. 0.041;
Matches 134;   Conservative   0;   Mismatches 123;   Indels   3;   Gaps   1;

Qy 1218 CGCGCCCGAGGCCGAGCCGGCCTTCATGCGGCCTATGCGGCTATGGCATCCGTTG 1277
          ||||| |||| ||||||| || || ||| |||||
Db   720 CGCCGCGCCTCGCCGAGCGCGTGCTGGGGCAAGTTCGTCAACGCGGCCAGACGTGCGT 779

Qy 1278 TTTGTCGGAATACGAGAGTTCGTCGTGGGGCGCTTTGATCGAGCTGCCTTCAAAAGCCCGGCCAA 1337
          || || ||| |||||| ||| |||
Db   780 GGCGGCGGACTACGTGTTCGTGCACGCGCGTCGTGGGGCGGCCGGCGGGTTCCTGGACGCGGCTCAA 839

Qy 1338 GCTGGCTATTGTGCCTTTGCAGGACGTGCTCGGGGCTGGGCCCGGGCATGAACTT 1397
          |||| || || || ||| ||| ||||
Db   840 GGGGTCCATCACGGCGCTTCTACGCGCCGGGCACCGAGCAGCAGGAGCCCCTGACCT 899

Qy 1398 CCCCGGACACGGCTCGGGGGACAACTGGCGTGGCCGCTACGCGAGGCGGACCTCGAGCCCCG 1457
          || |||||| ||||||||||

; Sequence 1116, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1116
; LENGTH: 15377
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1116

Query Match   7.5%;   Score 113.2;   DB 3;   Length 15377;
Best Local Similarity   53.8%;   Pred. No. 9.6e-19;
Matches 269;   Conservative   0;   Mismatches 213;   Indels 18;   Gaps   1;

Qy   399 CGAGGGCCGAGCCTTCTGGAAGACTATGCGCTCTTTATGGCCCTCAAGACCCGGTT 458
          ||||| ||||||||||||| || ||||| ||||| |||||
Db 8803 CGACGCCCAGGGCGAATGGCTGGAGAGCTACGGCGCTCTTCACCGCCATCAGCGAGCAGGA 8744

Qy   459 TGACGGCAAGCCCTGGAACAGAGTGGAGCCCCGAGCGTCGCGACCGTGACCCGCTGCCCT 518
          ||||| ||||| ||||||| |||
Db 8743 AGAGCCGCCGCCGTGGTGGAGTGCGGCCGAGCCGCGCTGCCACCCGTCAACCCGGAGCCCT 8684

Qy   519 GGCCAGGGCCCGTGAGGAGCTGGCCGAGGAGCTGGCCCTTTACGAGTGGGATTCAGTGGCT 578
          |||| |||||||||| | |||| |||
Db 8683 GGCCCAGAGTCCCGGAGCTGGACTGAGCTGGAGCCGCGCGTGTGAGCGCTGGCTCCAGTGGGT 8624

Qy   579 TTTTTATCTGGAATGGGGCCAGAGCCATGGGGCGAATCCAAGGGGATTCAGATTAT 638
          ||||||| ||||| |||||||||||
Db 8623 GGCCGAGCAGCAGTGAACAGAGGTGCGGAACAGACGGTCCCCGGTGCTCCTCCTGTG 8564

Qy   639 CGGCGATATGCCCATCTTTGTGCGCCTTCGATTCCTCAGATGTCTGGGCCAACCCGCAGTA 698
          |||||||| |||||||||
Db 8563 CGGGCGACGAGCCGTTCATCATCGGTCAGGACAGCAGCGCGGCGTGTGGGGCGCACCCGGACAT 8504

Qy   699 CTTCTACCTCGAGGCCGATGGCAACCCCACGGTGGTGCGGCGTTCCGCGGGGACTACTT 758
          ||||| || ||||||
Db 8503 CCTGCGGCGGCGACGCC------------CGCCTGGGCGTGCCTCCCGGATGCGACTT 8462

Qy   759 CTCCGAAACCGGCCAGCTCTGGGCAATCCGCTCTATCGCCTCTGGGATGTGATGGAAAGGGA 818
          |||||| ||| |||||||
Db 8461 CTCCGGCCACAGGGCCAGGACTGGGGCCTGTGCCCTCACCTTCGACCCTTCGGCGGCCATGGAGAGGA 8402

Qy   819 CAACTTTGCCTGTGTGCATTGCCCGCGATAAGGCAGTCGCTCAAGCAGTGCCACCTGGTGCG 878
          |||| |||||
Db 8401 CGACTTCGCGTGCGTGGCTGCTGAAGACGGCGCGCGAAGAAGGGCGCGCCAGCTACTACGACTTGCCCG 8342

Qy   879 CATCGACCACTTCCGCGGGT 898
          ||||| ||||| ||||||
Db 8341 CGTGGACCACGCGCGGTGGGT 8322

RESULT 7
US-09-719-978-1
; Sequence 1, Application US/09719978
; Patent No. 6858717
; GENERAL INFORMATION:
; APPLICANT: Biogemma
; TITLE OF INVENTION: Method for obtaining modified polysaccharides
; FILE REFERENCE: BFF 98/0267
; CURRENT APPLICATION NUMBER: US/09/719,978
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: FR/9807589
; PRIOR FILING DATE: 1998-06-16

```
Db    900  GACGCGA---CTGGTGGACCCGTGGCGCGCGGGTGAAGGACCTGCTCAAGCCGCTC 956

Qy   1458  TCTGGCCGCGGGACTGCGGG 1477
           ||||||||||||||
Db    957  GGTGGCCGCGGGAGCGAAGG 976
```

RESULT 9
US-09-902-540-1091
; Sequence 1091, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1091
; LENGTH: 12194
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(12194)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1091

```
Query Match      3.1%;     Score 47.2;  DB 3;  Length 12194;
Best Local Similarity   51.5%;    Pred. No. 0.068;
Matches  134;    Conservative   0;  Mismatches 123;  Indels   3;  Gaps   1;

Qy  1218  CGCGCCGGAGGCCGAGCGGCGGCCCTTCATGCGGCCTACCTGGCCCGCTATGGCATCCGTTG 1277
          |||||| ||||| | ||| |||  ||| ||||||| ||| || |||||| |
Db  2146  CGCCGCCGCTGCGGAGCGGCCTTGCGTGGGGCAAGTTCGTCAACGGCGACGTGCGT 2205

Qy  1278  TTTGTCGGAATACGAGTCGCGGCGCTTGATCGAGTCGGCCTTCAAAAGCCCGGCCAA 1337
          |||||| ||| |||||| ||||  ||| | ||| |  |||||||||||
Db  2206  GGGCGCCGGACTACGTGTTCGTGCACGCGTCGAAGAGCGGCGGCGTTCTGGACGCGCTCAA 2265

Qy  1338  GCTGGCTATTGTGCCTTTGACGAGGACGTGCTGTGGGCGTGG 1397
          | ||| |||  |||| || ||| || ||| | ||| ||
Db  2266  GGGCGTCCATCACGCGCTTCACGGCGGCACCGAGGAGCGGCGAGCCCTGACCT 2325

Qy  1398  CCCCGGACGGCTGGGGGACAACTGGGCGTTGGGCGCCACGCCCGAGGCGACCTGAGCCCGG 1457
          |||  ||| ||| ||  ||||  |||  | ||||  |||||| || |||||| |||
Db  2326  GACGCGA---CTGGTGGACCCGTGGCGCGCGGGTGAAGGACCTGCTCAAGCCGCTC 2382

Qy  1458  TCTGGCCGCGGGACTGCGGG 1477
          ||||||||||||||
Db  2383  GGTGGCCGCGGGAGCCAAGG 2402
```

RESULT 10
US-08-838-543-1
; Sequence 1, Application US/08838543
; Patent No. 5994623
; GENERAL INFORMATION:
; APPLICANT: KREBBERS, ENNO
; APPLICANT: BROGLIE, KAREN E.
; TITLE OF INVENTION: CORN 4-(-GLUCANOTRANSFERASE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DUPONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.50 INCH DISKETTE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD FOR WINDOWS 95 (VER. 7.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/838,543
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: MAJARIAN, WILLIAM R.
; REGISTRATION NUMBER: P-41,173
; REFERENCE/DOCKET NUMBER: BB-1101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4926
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1683 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1489
US-08-838-543-1

```
Query Match       3.1%;     Score 46.4;  DB 2;  Length 1683;
Best Local Similarity   53.3%;    Pred. No. 0.054;
Matches  98;    Conservative   0;  Mismatches   86;  Indels   0;  Gaps   0;

Qy   737  CGGGCCGTTCCGCGGGACTACTTCTCCGAAACCGGCCAGCTCTGGGCAATCCGCTCTATC 796
          |  ||||| |||| || ||| |||| | ||  |||| || || | |||  ||  |||  ||
Db   171  CTGGAGCGGCCTCCTGATTATTTTGACAAGAATGGACAAAATTGGGGTTTTCCTACATATA 230

Qy   797  GCTGGGATGTGATGGAAGAGGACAACTTGCCTCGGTGCATTGCCCGCCATAAGGCAGTCGC 856
          |||||  ||| || || || ||| |||| || ||||| | ||| || |||| ||  ||||
Db   231  ACTGGGAGGAGATGTCAAAGGATAATTATGGTGGTGGCGAGCTCGTCTGACACAGATGG 290

Qy   857  TCAAGCAGTGCCACCTGGTGCGCATCGACCACTTCCCGCGGGTTTGAGCGCTACTGGGAGG 916
          |  ||| || ||||| || || |||| | |||| ||  | || |  |||| || ||||
Db   291  CAAAGTACTTCACAGCATACAGGATAGACCACATCTTGGGTTTCTTTAGGATATGGGAGC 350

Qy   917  TTCC 920
          ||||
Db   351  TTCC 354
```

RESULT 11
US-09-780-173A-10/c
; Sequence 10, Application US/09780173A
; Patent No. 6455307
; GENERAL INFORMATION:
; APPLICANT: Robert McKay
; APPLICANT: Susan M. Freier
; APPLICANT: Jacqueline Wyatt
; TITLE OF INVENTION: ANTISENSE MODULATION OF CASEIN KINASE 2-ALPHA PRIME EXPRESSION
; FILE REFERENCE: RTS-0165
; CURRENT APPLICATION NUMBER: US/09/780,173A
; CURRENT FILING DATE: 2001-02-08
; NUMBER OF SEQ ID NOS: 95
; SEQ ID NO 10
; LENGTH: 1877
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (344)...(1396)
US-09-780-173A-10

```
Query Match        3.1%;     Score 46.4;   DB 3;  Length 1877;
```

Best Local Similarity   53.3%;   Pred. No. 0.057;   Indels   0;   Gaps   0;
Matches   98;   Conservative   0;   Mismatches   86;   Indels   0;

```
Qy   1312  GAGCTGGCCTTCAAAAGCCCGGCCAAGCTGGCTATTGTGCCTTTGCCTTGCCTGCTGGGG  1371
                 ||   ||||     ||   |   |||||   ||||   |||   |
Db    436  GACGTGGGCTTCGTAGTCCCAGTACTCGCGGCTCCTCAGCGTGTTCACCTCGGCGTAGAC   377

Qy   1372  CTGGGCCCCGAGGCCCGCATGAACTTCCCCCGACGCGCTGGGGGACAACTGGGCGTCGGCGC  1431
                 |||||||       |||        ||||      ||
Db    376  CCGGGGCCCGACTCCCGGCGGCGCGGGGGCGCGGCCGGGACCCGGGGCGGCGCCCGGG   317

Qy   1432  TACGCCGAAGGCGACCTCGAGCCCGTCTGCGCCCGCGGGGACTGCGGCCCTTGGCCCGAGGCC  1491
                 |||||   |||||     ||
Db    316  GGGGCCGAGGAGGCGCGGCGGCCGGCGGGGCGCGGGGGCGCGCAGCGCAGCGGAAGCGGCG   257

Qy   1492  AGCC  1495
           ||||
Db    256  GGCC   253
```

RESULT 12
US-09-105-537-32
; Sequence 32, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;   APPLICANT: Sherman, D.H.
;   APPLICANT: Liu, H.
;   APPLICANT: Xue, Y.
;   APPLICANT: Zhao, L.
;   TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;   FILE REFERENCE: 600.438US1
;   CURRENT APPLICATION NUMBER: US/09/105,537A
;   CURRENT FILING DATE: 1998-06-26
;   NUMBER OF SEQ ID NOS: 43
;   SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-32

Query Match   3.1%;   Score 46.2;   DB 3;   Length 11220;
Best Local Similarity   47.4%;   Pred. No. 0.12;   Indels   0;   Gaps   0;
Matches   138;   Conservative   0;   Mismatches   153;   Indels   0;

```
Qy   1211  TCCGCACCGCGCCGGAGGCCCGAGCGGCCTTCATGCGGGCCTTACCTGGCCGCCTATGGCA  1270
           |||    ||   |   ||||      ||   ||   ||   |   ||||||
Db   9737  TGCGCCGGGCCCGGACCGCGCCTTCCGGACGCGTTCGGCGGGCCCAGCGCCACACGGGCAAGGTCG  9796

Qy   1271  TCCGTTGTTGTTCGGAATACGAGGTCGGGGCCTTTGATCGAGCTGGCCTTCAAAAGCC  1330
                |     |||   ||   ||||   ||||        |||
Db   9797  TCCTCACGATGCCGTCGGGCCTCGACCCGGCTGCTGACCGGGCTACGGTACGGTCGTGGG  9856

Qy   1331  CGGCCAAGCTGGCTATTGTGCCTGGGGGACAACTGGGCGTGGCGTCGGCGCTCGCCGAGGGCCGCCA  1390
                 |||||        |||    ||||     ||       |
Db   9857  GTGCGCTGGGGGGCATCGTGGTGGGCGAGTGGGGCGTACGCGCGCCGGGCCTGC  9916

Qy   1391  TGAACTTCCCCGACGGCTGGGGGGACAACTGGGCGCGTCGCGGCCTACGCCGAAGGCGACCTCG  1450
                 |||||         ||     ||||||        |||
Db   9917  TGCTCGTGAGCCGGCGGGGCACGGACGCCCCGGGCCGCGGGCCGCAGCTCGTGCACGAGCTGG  9976

Qy   1451  AGCCCGGTCTGCCCGCGGGGACTGCGGGCCCTGCGCGAGGGCCAGCCAGCGCG  1501
                |||||       ||  |||||    ||
Db   9977  AGGCCCTGGGAGCCCGACGTCTCGGTGGCCGTGCGGTGCGGGCCGACCGGCG  10027
```

RESULT 13
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;   APPLICANT: Sherman, D.H.
;   APPLICANT: Liu, H.

;   APPLICANT: Xue, Y.
;   APPLICANT: Zhao, L.
;   TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;   FILE REFERENCE: 600.438US1
;   CURRENT APPLICATION NUMBER: US/09/105,537A
;   CURRENT FILING DATE: 1998-06-26
;   NUMBER OF SEQ ID NOS: 43
;   SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-5

Query Match   3.1%;   Score 46.2;   DB 3;   Length 36778;
Best Local Similarity   47.4%;   Pred. No. 0.18;   Indels   0;   Gaps   0;
Matches   138;   Conservative   0;   Mismatches   153;   Indels   0;

```
Qy    1211  TCCGCACCGCGCCGGAGGCCCGAGCGGCCTTCATGCGGGCCTTACCTGGCCGCCTATGGCA  1270
            |||    ||   |   ||||      ||   ||   ||   |   ||||||
Db   25424  TGCGCCGGGCCCGGACCGCGCCTTCCGGACGCGTTCAGCCAGCCCGCCACACGGGCAAGGTCG  25483

Qy    1271  TCCGTTGTTGTTCGGAATACGAGGTCGGGGCCTTTGATCGAGCTGGCCTTCAAAAGCC  1330
                 |     |||   ||   ||||   ||||        |||
Db   25484  TCCTCACGATGCCGTCGGGCCTCGACCCGGCTGCTGACCGGGCTACGGTCGTGGG  25543

Qy    1331  CGGCCAAGCTGGCTATTGTGCCTTTGCCAGGACGTGCGTTGGGGTCGGGGCCCCGAGGCGCCGCCA  1390
                 |||||        |||||      ||||      ||       |
Db   25544  GTGCGCTGGGGGGCATCGTGGTGGGGCGACGTGGTGGGCGAGTGGGGCGTACGACGCCTGC  25603

Qy    1391  TGAACTTCCCCGACGGCTGGGGGACAACTGGGCGTGCGCGTGGCGTACGCCGAAGGCGACCTCG  1450
                 |||||         ||     ||||||        |||
Db   25604  TGCTCGTGAGCCGGCGGGGCACGGACGCCCCGGGCCGCGGGCCGAGCTCGTGCACGAGCTGG  25663

Qy    1451  AGCCCGGTCTGCCCGCGGGGACTGCGGGCCCTGCGCGAGGGCCAGCCAGCGCG  1501
                |||||       ||  |||||    ||
Db   25664  AGGCCCTGGGAGCCCGACGTCTCGGTGGCCGTGCGGTGCGGGCCGACCGGCG  25714
```

RESULT 14
US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
;   APPLICANT: ASHLEY, Gary
;   APPLICANT: BETLACH, Melanie C.
;   APPLICANT: BETLACH, Mary C.
;   APPLICANT: McDANIEL, Robert
;   APPLICANT: TANG, Li
;   TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;   FILE REFERENCE: 30062002120
;   CURRENT APPLICATION NUMBER: US/09/320,878A
;   CURRENT FILING DATE: 1999-05-27
;   EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;   EARLIER FILING DATE: 1998-08-28
;   EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;   EARLIER FILING DATE: 1998-05-06
;   EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;   EARLIER FILING DATE: 1997-04-30
;   EARLIER APPLICATION NUMBER: 60/119,139
;   EARLIER FILING DATE: 1999-02-08
;   EARLIER APPLICATION NUMBER: 60/100,880
;   EARLIER FILING DATE: 1998-09-22
;   EARLIER APPLICATION NUMBER: 60/087,080
;   EARLIER FILING DATE: 1998-05-28
;   NUMBER OF SEQ ID NOS: 34
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 38506
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-19

Query Match   3.1%;   Score 46.2;   DB 3;   Length 38506;
Best Local Similarity   47.4%;   Pred. No. 0.18;
Matches 138;   Conservative   0;   Mismatches 153;   Indels   0;   Gaps   0;

Qy   1211 TCCGCACCGCGCCGGAGGCCGAGCGGCCTTCATGCGGGCCTACCTGGCCCGCTATGGCA 1270
             | ||||||||| | ||||| |||| ||| ||| |||||||||||||| |||||
Db   23566 TGCGCCGGCGCCCCGACGCGCCTTCGGCGACGTCAGCCAGGCCCGCCACACGGGCAAGGTCG 23625

Qy   1271 TCCGTTGTTTGTCGGAATACGAGTCGGCGGGGCCTTTGATCGAGCTGGCCTTCAAAAGCC 1330
             |||| || | |||| ||| | ||| | ||| || |||| ||||||| ||||||
Db   23626 TCCTCACGATGCCGTCGGCGCTCGACCCGGAGGTACGGTCCTGCTGACCGGCGGCACCG 23685

Qy   1331 CGGCCAAGCTGGCTATTGTGCCTTTGCAGGACGTGCTGGGCGTGGGCCCCGAGGCCCGCA 1390
             ||| || |||| | ||||||||||| | |||| ||| |||| ||||||||| |||||
Db   23686 GTGCGCTGGGGGGGCATCGGCCCCGGCACGTGGTGGGGCGAGTGGGGCGTACGACGCCTGC 23745

Qy   1391 TGAACTTCCCCGACGGCTGGGGGACAACTGGCGCGTGGGCGTACGCCGAAGGCGACCTCG 1450
             ||||||||||| ||| | ||| | | |||| | |||| |||||||||||| ||||||
Db   23746 TGCTCGTGAGCCATGCTCGGCGCCTGGGGGGGCACCGACGCCGCCCGGCGAGCTCGTGCACGAGCTGG 23805

Qy   1451 AGCCCGGTCTCGGCCGCGGGACTGCGGCGCCTGCCGAGGCCAGCCAGCGCG 1501
             |||||||||| | ||| | || |||| ||||||| |||||||||||
Db   23806 AGGCCCTGGGAGCCGACGTCTCGGTGGCCGGGGCGTGCGACGTCGCCGACCGCG 23856

Search completed: January 14, 2006, 14:49:35
Job time : 303 secs

RESULT 15
US-09-141-908-1
; Sequence 1, Application US/09141908
; Patent No. 6503741
; GENERAL INFORMATION:
;   APPLICANT: ASHLEY, Gary
;   APPLICANT: BETLACH, Melanie C.
;   APPLICANT: BETLACH, Mary
;   APPLICANT: MCDANIEL, Robert
;   APPLICANT: TANG, Li
;   TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
;   TITLE OF INVENTION: Modular PKS Gene Cluster as Scaffold
;   FILE REFERENCE: 30062002100
;   CURRENT APPLICATION NUMBER: US/09/141,908
;   CURRENT FILING DATE: 1998-08-28
;   EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;   EARLIER FILING DATE: 1998-05-06
;   EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;   EARLIER FILING DATE: 1997-04-30
;   EARLIER APPLICATION NUMBER: PROV. 60/076,919
;   EARLIER FILING DATE: 1998-03-05
;   EARLIER APPLICATION NUMBER: PROV. 60/087,080
;   EARLIER FILING DATE: 1998-05-28
;   NUMBER OF SEQ ID NOS: 31
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 38506
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-141-908-1

Query Match   3.1%;   Score 46.2;   DB 3;   Length 38506;
Best Local Similarity   47.4%;   Pred. No. 0.18;
Matches 138;   Conservative   0;   Mismatches 153;   Indels   0;   Gaps   0;

Qy   1211 TCCGCACCGCGCCGGAGGCCGAGCGGCCTTCATGCGGGCCTACCTGGCCCGCTATGGCA 1270
             | ||||||||| | ||||| |||| ||| ||| |||||||||||||| |||||
Db   23566 TGCGCCGGCGCCCCGACGCGCCTTCGGCGACGTCAGCCAGGCCCGCCACACGGGCAAGGTCG 23625

Qy   1271 TCCGTTGTTTGTCGGAATACGAGTCGGCGGGGCCTTTGATCGAGCTGGCCTTCAAAAGCC 1330
             |||| || | |||| ||| | ||| | ||| || |||| ||||||| ||||||
Db   23626 TCCTCACGATGCCGTCGGCGCTCGACCCGGAGGTACGGTCCTGCTGACCGGCGGCACCG 23685

Qy   1331 CGGCCAAGCTGGCTATTGTGCCTTTGCAGGACGTGCTGGGCGTGGGCCCCGAGGCCCGCA 1390
             ||| || |||| | ||||||||||| | |||| ||| |||| ||||||||| |||||
Db   23686 GTGCGCTGGGGGGGCATCGGCCCCGGCACGTGGTGGGGCGAGTGGGGCGTACGACGCCTGC 23745

Qy   1391 TGAACTTCCCCGACGGCTGGGGGACAACTGGCGCGTGGGCGTACGCCGAAGGCGACCTCG 1450

```
                 GenCore version 5.1.6
          Copyright (c) 1993 - 2006  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:      January 14, 2006, 13:11:49 ; Search time 1145 Seconds
                                (without alignments)
                    10854.913 Million cell updates/sec

Title:       US-10-618-976-1
Perfect score: 1503
Sequence:    1 atgcaactccaacgcgcctt.........ccgaggccagccagccgct 1503

Scoring table: IDENTITY_NUC
             Gapop 10.0 , Gapext 1.0

Searched:    9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters:     19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Published Applications NA_Main:*
             1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
             2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
             3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
             4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
             5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
             6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
             7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
             8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
             9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
            10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

                          SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

Result No.	Score	% Match	Query Match	Length	DB	ID	Description
1	1503	100.0	1503	1503	9	US-10-618-976-1	Sequence 1, Appli
2	287.4	19.1	1515	1515	8	US-10-472-928-4401	Sequence 4401, Ap
3	287.4	19.1	5910	5910	2	US-08-961-527-165	Sequence 165, App
4	287.4	19.1	5910	5910	7	US-10-158-844-165	Sequence 165, App
5	287.4	19.1	2162598	2162598	8	US-10-472-928-4979	Sequence 4979, Ap
6	265.6	17.7	1377	1377	9	US-10-617-320-336	Sequence 336, App
7	159	10.6	1676	1676	8	US-10-425-114-31307	Sequence 31307, A
8	134	8.9	1788	1788	8	US-10-425-115-140665	Sequence 140665,
9	104.2	6.9	540	540	7	US-10-767-701-19455	Sequence 19455, A
10	102.2	6.8	886	886	7	US-10-437-963-36082	Sequence 36082, A
11	99.2	6.6	2540	2540	7	US-10-424-599-28136	Sequence 28136, A
12	84.2	5.6	2220	2220	6	US-10-156-761-5375	Sequence 5375, Ap
13	84.2	5.6	9025608	9025608	6	US-10-156-761-1	Sequence 1, Appli
14	76.8	5.1	563	563	7	US-10-424-599-28135	Sequence 28135, A
15	66.8	4.4	1945	1945	7	US-10-425-114-29452	Sequence 29452, A
16	64.6	4.3	259	259	3	US-09-923-876-1014	Sequence 1014, Ap
17	64.6	4.3	259	259	3	US-09-923-876-1014	Sequence 1014, Ap
18	56.8	3.8	2217	2217	3	US-09-938-842A-249	Sequence 249, App
19	56.8	3.8	2217	2217	3	US-09-938-842A-249	Sequence 249, App
20	56.2	3.7	459	459	6	US-10-091-007-115	Sequence 115, App
21	55.6	3.7	978	978	6	US-10-107-431-150	Sequence 150, App
22	55.6	3.7	45055	45055	6	US-10-107-431-277	Sequence 277, App
23	52.8	3.5	3771	3771	6	US-10-156-761-5283	Sequence 5283, Ap
24	50	3.3	36401		8	US-10-925-357-1	Sequence 1, Appli
25	49.4	3.3	1959		8	US-10-739-930-3025	Sequence 3025, Ap
26	48.6	3.2	745		6	US-10-336-597-7	Sequence 7, Appli
27	48.6	3.2	2241		6	US-10-156-761-1111	Sequence 1111, Ap
28	48.2	3.2	1146		6	US-10-163-198-12	Sequence 12, Appl
29	47.6	3.2	2658		7	US-10-282-122A-25660	Sequence 25660, A
30	47.4	3.2	1836		8	US-09-738-626-2483	Sequence 2483, Ap
31	47.4	3.2	1959		8	US-10-781-014-381	Sequence 381, App
32	47.4	3.2	2056		6	US-10-494-836-45	Sequence 45, Appl
33	47.4	3.2	2351		7	US-10-437-963-499	Sequence 499, App
34	47.4	3.2	3309400		3	US-09-738-626-1	Sequence 1, Appli
35	47.2	3.1	294		8	US-10-425-115-50318	Sequence 50318, A
36	47.2	3.1	1425		6	US-10-369-491-43028	Sequence 43028, A
37	47	3.1	1191		7	US-10-282-122A-14804	Sequence 14804, A
38	46.8	3.1	1102		7	US-10-425-114-24094	Sequence 24094, A
39	46.8	3.1	1123		8	US-10-425-115-103771	Sequence 103771,
40	46.6	3.1	1545		6	US-10-163-198-96	Sequence 96, Appl
41	46.6	3.1	1947		9	US-10-928-992-25	Sequence 25, Appl
42	46.6	3.1	1948		7	US-10-260-238-298	Sequence 298, App
43	46.4	3.1	930		6	US-10-156-761-2723	Sequence 2723, Ap
44	46.4	3.1	1010		8	US-10-425-115-173269	Sequence 173269,
45	46.4	3.1	1683		6	US-10-336-597-1	Sequence 1, Appli

```
                          ALIGNMENTS

RESULT 1
US-10-618-976-1
; Sequence 1, Application US/10618976
; Publication No. US2005074769A1
; GENERAL INFORMATION:
;  APPLICANT: Fukuyama, Shiro
;  TITLE OF INVENTION: Polypeptides Having Glucanotransferase
;  TITLE OF INVENTION: Activity And Nucleic Acids Encoding Same
;  FILE REFERENCE: 6012.200-US
;  CURRENT APPLICATION NUMBER: US/10/618,976
;  CURRENT FILING DATE:   2003-07-14
;  PRIOR APPLICATION NUMBER: US/09/687,360
;  PRIOR FILING DATE: 2000-10-13
;  PRIOR APPLICATION NUMBER: PA 1999 01501
;  PRIOR FILING DATE: 1999-10-20
;  PRIOR APPLICATION NUMBER: PA 1999 01641
;  PRIOR FILING DATE: 1999-11-15
;  PRIOR APPLICATION NUMBER: 60/160,903
;  PRIOR FILING DATE: 1999-10-22
;  PRIOR APPLICATION NUMBER: 60/166,539
;  PRIOR FILING DATE: 1999-11-18
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;  LENGTH: 1503
;  TYPE: DNA
;  ORGANISM: Thermus rubens
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (1)...(1503)
US-10-618-976-1

Query Match             100.0%;  Score 1503;  DB 9;  Length 1503;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 1503;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1 ATGCAACTCCAACGCGCTTTGGGAATTTGCTCCACCCCACCCAGTTTCCGGGTCGCTGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGCAACTCCAACGCGCTTTGGGAATTTGCTCCACCCCACCCAGTTTCCGGGTCGCTGG 60

Qy     61 GGGATTGGGCTCTGGGCCGCGAGGCCGAGCGGTTTTTGGACTGGCTGGCCGATGCGGGA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GGGATTGGGCTCTGGGCCGCGAGGCCGAGCGGTTTTTGGACTGGCTGGCCGATGCGGGA 120

Qy    121 GCCCGCTGGTGGCAGTCTTACCGCTGGGCCCTACCGCTGGGCCACTCGGCGACTCCCGTACCAG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
RESULT 2
US-10-472-928-4401
; Sequence 4401, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
;  TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE: P026926WO
;  CURRENT APPLICATION NUMBER: US/10/472,928
;  CURRENT FILING DATE: 2003-09-26
;  PRIOR APPLICATION NUMBER: GB-0107658.7
;  PRIOR FILING DATE: 2001-03-27
;  NUMBER OF SEQ ID NOS: 4979
;  SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 4401
;  LENGTH: 1515
;  TYPE: DNA
;  ORGANISM: Streptococcus pneumoniae
US-10-472-928-4401

  Query Match             19.1%;  Score 287.4;  DB 8;  Length 1515;
  Best Local Similarity   51.7%;  Pred. No. 7e-70;
  Matches  763;  Conservative    0;  Mismatches  696;  Indels  18;  Gaps   4;
```

```
Qy      11  AACGGCGCTTTGAATTTTGCTCACCCCACCAGTTTCCGGGTCGCTGCGGGGATTGGGG   70
            |||  || || ||||   ||| || | ||||||| ||  |||  ||| ||| |||||
Db       8  AACGTCAAAGTGGTGTGTTGATGCACATCTCTTCTTCCAGGAGCTTACGGAATCGGAT   67

Qy      71  CTCTGGGCGCCGCGAGGCCCGAGGCCGGTTTTTGGACTGGCTGGCCGATGCGGGAGCCCGCTGCT  130
            ||  ||   ||  || ||    |||||||  ||      |||||||  ||    ||||| | | |
Db      68  CATTTGGTCAAAAGTGCTTACGACTTCGTTGATTTCTTGGTCCGTCCGTACAAAACAACGTTACT  127

Qy     131  GGCCAGGTCTTACCGCTGGGCGTCGCCGGGCGATGCGTTACGGGACTTACGGACTTCCTTCTCGG  190
            ||||  ||| |  |  ||| || || ||  ||    || |  ||| ||  ||   || | |||
Db     128  GGCAAATCCTTCCATTAGGAGCACTAGTTACGGGATTCTTACCGGATTCTTCCAAATCTTTCTCAG  187

Qy     191  CTTTTGCCGGTAACCCGTATTTGGTTGACCCGGAGATGCTGATTGAAAAGGCTGGCTGG  250
            ||||||| || |    |  || | ||||  ||||  || | |  | |||| |||  |||
Db     188  CCTTCGCAGGAAACACTCATTTTATCGATTTAGTATCTTGGTTGGAGCAAGGTTTGTTGG  247

Qy     251  AACAAAGGGAAGCGCCCCCGCCTGTTGCGGGCCTTGCGGGCGTTTCGGCAGGGCTTCGGCGC  310
            || ||||  |||||  ||| | |  || |  || ||   ||   ||  || | | ||| ||||
Db     248  AAGCAAGTGACCTTGAAGAAGTGACTTGGTAGCGATGCGATGCGTCTGAAGTTGACTATGCTA  307

Qy     311  AGACCCCGCTGGCCCTGGCGCCCTGTTGCGGCGGGGCTTTCGGCCAAGGGCTTCGGCCC  370
            ||||| |||||  |  |  |||| || | |  || |  ||||||  ||| | | ||  ||
Db     308  AAATCTACTATGCCACGTGCTCTTTTTAGAAAAAGCCGTGAAAAGCGGTTCTTTGAAGTCG  367

Qy     371  AGGATAAGACCCGACTGGAAGCCTTATCGAGCCATATG  430
            ||     |  |     ||| |  || |  |||
Db     ...
```

```
Db     121  GCCCGCTGGTGCCAGGTCTTACCGCTGGGCCTACCAGTTACGGGGACTGCGGCGTACCAG  180
Qy     181  TCCTTCTCGGCGTTTGCCGGTAACCCGTATTTGGTTGACCCCGAGATGCTGATTGAAAAA  240
            |||||||||||||| |||||||||||| ||| |  ||||| |||||| || ||||||||||
Db     181  TCCTTCTCGGCGTTTGCCGGTAACCCGTATTTGGTTGACCCCGAGATGCTGATTGAAAAA  240
Qy     241  GGCTGGCTGGAACAAAGCGAAGCGCCCCGCCTATCGCCAGCGCGTGGAT TATGGC  300
            ||||||||||||||||| |||||||||||  ||||||| |  | ||||  |||||||||
Db     241  GGCTGGCTGGAACAAAGCGAAGCGCCCCGCCTATCGCCAGCGCGTGGATTATGGC  300
Qy     301  TGGCTTTACCAGACCCCTGGCCCCTGGTTGCGGGGGTTTCGGCGCAAGG  360
            |||||||||||||||||||||||||||||||||||||||| ||||||||
Db     301  TGGCTTTACCAGACCCCTGGCCCCTGGTTGCGGGGGTTTCGGCGCAAGG  360
Qy     361  GCTTCGGCCCCAGGATAAGACCCGACTGGAAGGCCTTTATCGAGGCGCAAGCGCCTTCTGGCTG  420
            |||||||||||||||||||||||||||||||||||||| |||||| || || || ||| |||||
Db     361  GCTTCGGCCCCAGGATAAGACCCGACTGGAAGGCCTTTATCGAGGCGCAAGCGCCTTCTGGCTG  420
Qy     421  GAAGACTATGCGGCTCTTTATGGCCCTTCAAGACCCGGTTGACGGCAAGCCCGGAACGAG  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421  GAAGACTATGCGGCTCTTTATGGCCCTTCAAGACCCGGTTGACGGCAAGCCCGGAACGAG  480
Qy     481  TGGAGCCCCGACGCTGCGCGACCGGCTGCCCTGGCCAGGGCCGTGAGGAGCTG  540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481  TGGAGCCCCGACGCTGCGCGACCGGCTGCCCTGGCCAGGGCCGTGAGGAGCTG  540
Qy     541  GCCGAGGAGGTGGCCCTTTACGATGGCTTTTTATCTGGAATGGGGCCAG  600
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db     541  GCCGAGGAGGTGGCCCTTTACGATGGCTTTTTATCTGGAATGGGGCCAG  600
Qy     601  ACCAAGGCCTATGCCGAATCCAAGGGATTCAGATTATCGGCGATATGCCCATCTTTGTG  660
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601  ACCAAGGCCTATGCCGAATCCAAGGGATTCAGATTATCGGCGATATGCCCATCTTTGTG  660
Qy     661  GCCTTCGATTCCTCAGATGTCTGGGCCAACCCGAGTACTTCTACCTCGAGGCCGATGGC  720
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661  GCCTTCGATTCCTCAGATGTCTGGGCCAACCCGAGTACTTCTACCTCGAGGCCGATGGC  720
Qy     721  AACCCACGGTGGTGCGGGCGTTCCGCGGGACTACTTCTCGCAGAACCGGCCAGCTCTGG  780
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721  AACCCACGGTGGTGCGGGCGTTCCGCGGGACTACTTCTCGCAGAACCGGCCAGCTCTGG  780
Qy     781  GGCAATCCGCTCTATCGCTCTGGGATGTGATGGAAAGGGACAACTTTGCCTGGTGCATTGCC  840
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781  GGCAATCCGCTCTATCGCTCTGGGATGTGATGGAAAGGGACAACTTTGCCTGGTGCATTGCC  840
Qy     841  CGCATAAGGCAGTCGCTCAAGCAGTGCCACCTGTGCGCCATCGACCACTTCCGGCGGTTT  900
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841  CGCATAAGGCAGTCGCTCAAGCAGTGCCACCTGTGCGCCATCGACCACTTCCGGCGGTTT  900
Qy     901  GAAGCCTACTGGGAGGTTCCGTTTGGCGGGGCGCGCCAATGCTGTGAGGGGCGCGTGGGTCAAA  960
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901  GAAGCCTACTGGGAGGTTCCGTTTGGCGGGGCGCGCCAATGCTGTGAGGGGCGCGTGGGTCAAA  960
Qy     961  GCCCCCAGGGGAGAAGCTGTTTGCTGCGCGTCGCCGGGCGATGCGCCCCATCATT  1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961  GCCCCCAGGGGAGAAGCTGTTTGCTGCGCGTCGCCGGGCGATGCGCCCCATCATT  1020
Qy    1021  GCCGAAGACCTGGGGTGATCACCCCGAGGTGGCTTGCGGGATGGCTTCGGGTTC  1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021  GCCGAAGACCTGGGGTGATCACCCCGAGGTGGCTTGCGGGATGGCTTCGGGTTC  1080
Qy    1081  CCCGGGCATGAAGATTTTGCAGTTTGCTTTTTCCGGTGAGGACAACGCCTTTTTTGCCCCAC  1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081  CCCGGGCATGAAGATTTTGCAGTTTGCTTTTTCCGGTGAGGACAACGCCTTTTTTGCCCCAC  1140
Qy    1141  AACTACCCCGCCACGCACGCGGGCCTTCATGCGGGCCTTACCTGCC  1200
            ||||||||||||||||||||||||||||||||||||||||||||||
Db    1141  AACTACCCCGCCACGCACGCGGGCCTTCATGCGGGCCTTACCTGCC  1200
Qy    1201  CTGGGGATGGTTCCGCACCGGCCGAGGCCGCGAGCCGCCTTCGGCGCC  1260
            |||||||||||||||||||||||||||||||||||||||||||||||
Db    1201  CTGGGGATGGTTCCGCACCGGCCGAGGCCGCGAGCCGCCTTCATGCCC  1260
Qy    1261  CGCTATGGCATCCGTGTGTTTGTCGGAATACGAGGTCGCGGGCGGCTGCC  1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261  CGCTATGGCATCCGTGTGTTTGTCGGAATACGAGGTCGCGGGCGGCTGCC  1320
Qy    1321  TTCAAAAGCCCGGCCAAGCTGGCTATTGTGCCTATTGTGCCGAGGACGTGCGGGGCCCC  1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321  TTCAAAAGCCCGGCCAAGCTGGCTATTGTGCCTATTGTGCCGAGGACGTGCGGGGCCCC  1380
Qy    1381  GAGGCCCGCATGAACTTCCCCGACAACTGGGCGTGGCCGGCTACGCCCGAA  1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381  GAGGCCCGCATGAACTTCCCCGACAACTGGGCGTGGCCGGCTACGCCCGAA  1440
Qy    1441  GGCGACCTCGAGCCCGTCTGCGGCCCGGAGGCCAGCCAGCGC  1500
            ||||||||||||||||||||||||||||||||||||||||||
Db    1441  GGCGACCTCGAGCCCGTCTGCGGCCCGGAGGCCAGCCAGCGC  1500
Qy    1501  GCT  1503
            |||
Db    1501  GCT  1503
```

RESULT 3
US-08-961-527-165
; Sequence 165, Application US/08961527
; Publication No. US20020032323A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Human Genome Sciences, Inc.
;   STREET: 9410 Key West Avenue
;   CITY: Rockville
;   STATE: Maryland
;   COUNTRY: USA
;   ZIP: 20850
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;   COMPUTER: HP Vectra 486/33
;   OPERATING SYSTEM: MSDOS version 6.2
;   SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/961,527
;   FILING DATE:
;   CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Brookes, A. Anders
;   REGISTRATION NUMBER: 36,373
;   REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (301) 309-8504
;   TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 165:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 5910 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
US-08-961-527-165

Query Match   19.1%;   Score 287.4; DB 2; Length 5910;
Best Local Similarity 51.7%;   Pred. No. 8.7e-70;
Matches 763; Conservative   0; Mismatches 696; Indels 18; Gaps   4;

Qy   11 AACGGCGCTTTGGAATTTGCTCCACCCCACCAGTTTTCCGGGTCGCTGGGGGATTCGGGG 70
              |||| ||||||||| | ||| ||||| ||||||||||||||||||||||||
Db   108 AACGTCAAAGTGGTGTTGGTGTGCAATCTCTTCTTCCAGGAGCTTACGGAGGAATTCGGAT 167

Qy   71 CTCTGGGCCGCGAGGCGCGAGCGGCGGTTTTGGACTGGCTGGCCGATGCGGGAGCCCGCTGGT 130
              | |||| ||| |||||| || | || |||||||||||||||||||
Db   168 CATTTGGTCAAAGTGCTTGGCCCTACGCATCCGTTCGTGATTTCTTGGTCCGTACAAAACAACGTTACT 227

Qy   131 GGCAGGTCTTACCGTGGGCCTCCGGACTCGGCCGACTCCAGTGCCTTCTCGG 190
              |||||||||||||| ||||||||||||||||||
Db   228 GGCAAATCCTTCCATTAGGAGCAACTAGTTACGGGGATTCTCCTTACCAAATCTTTCCTCAG 287

Qy   191 CTTTTGCCGGTAACCCGTATTTGGTGACCCCGAGATGCTGATTGAAAAAGGCTGGCTCG 250
              || |||||||||||||||
Db   288 CCTTCGCAGGAGAAACACTCATTTATGATTTAGATATCTTGGGAGGAGCAAGGTTTGTGTGG 347

Qy   251 AACAAAGCGAAGGCGCCCCCCGTCATCGGCCGTATCGGGCCGTGGATTATGGCCTTTACC 310
              |||||||||||||
Db   348 AAGCAAGTGACCTTGAAGGAGTTGACTTTGGTAGCGATGCGTCTGACTATGCTA 407

Qy   311 AGACCCGCTGCCCCCGTGTTGCGGCGGGTTTCGGCAAGGGGACC 370
              |||||||||
Db   408 AAATCTACTATGCACGTCGTCCTCTTTTGAGAAAAGCGGTGAAACGTTTCTTTGAAGACTCG 467

Qy   371 AGGATAAGACCCGACTGGAAGGCCTTTATCGAGGCCGCAGCGCTTCTCGCTGGAAGACTATG 430

Db   368 GAGATGTTAAAGATTTGCTCAAGACAACCAATCATGCTTGAGCTCTTG 427
           |||||
Qy   431 CGCTCTTTATGGCCGTCAAGACCCGGTTGACGCGCAAGCGAGTCG--AGCC 487
           | |||||
Db   428 CTGAGTATATGGCTATCAAAAGTATTTGACAAATCTTGCTTGGACTGAATGGCCAGATG 487

Qy   488 CCGAGCTGCGCGACCGTGAACCGGTCGCCTGCCCTGCAGGAGCGCCCTGGCCGAGG 547
           || ||||||
Db   488 CAGATGCTCGTCGTCCTGCTGGTAAAGCTTCAGCAGCTTGAAAGCTATCGTGAGCAATTGGCAGACA 547

Qy   548 AGGTGGCCCTTACGAGTGGATTCAGTGGCTTTTTATCTGGAAATGGGGCCAGACCAAGG 607
           ||||| |||||||
Db   548 AGTTGGTTTACCACCGTGGACTCAATACTTCTTCTTCCAACAATGGTTGAAATTGAAAG 607

Qy   608 CCTATGCCGAATCCAAGGGGATTCAGATATCGCGATATGCCCATCTTTGTGGCCTTCG 667
           |||||||||||
Db   608 CTTACGTAAGCGACAACCACATCGAAATCGTTGGGGACATGCCAATCTACGTAGCCGAAG 667

Qy   668 ATTCCTCAGATGTCTGGGCCAACCCGCAGTACTTCTACCTCGAGGCCGATGGCAACCCCA 727
           ||||||||| |||||||| |||||||
Db   668 ATTCAAGTGATATGTGGGCAAATCCACCATCTCTTCAAAAACAGATGTCAATGTAAGGCTA 727

Qy   728 CGGTGGTGGCGGGCGGGCGTTCCGGCGGGACTACTTCTCCGAAACCGGCAGCTCTGGGGCAATC 787
           ||||||||||||
Db   728 CTTGTATCGCAGGATGCCCACCAGATGAGTTTTCTGTAACTGGTGTCAGCTTTGGGGTAATC 787

Qy   788 CGCTCTATCGCTGGGATGTGATGGAAAGGGACAACTTGCCTGGTGCATTGCCCCGCATAA 847
           |||||||||| || |||| |||
Db   788 CAATCTATGACTGGGAGCATTGGAAGCAATGGACAAAGACGCTACAAATGGTGATTGAACGTTGC 847

Qy   848 GGCAGTCGCTCAAGCAGTGCCACTCGTGCGCATCGACCACTTCCGCGGGGTTTGAAGCCT 907
           ||||||| |||||||
Db   848 GTGAAAGCTTCAAAATCTCAAAATCTGTATCGCATCGGTGCCTTCGCGGCTTCGAATCTT 907

Qy   908 ACTGGGAGGTTCCGTTTTGGCCGGCCCAATGCGTGTGAGGGGCGCTGGTCAAAGCCCCAG 967
           |||||||| |||||
Db   908 ACTGGGAAATCCCTGCTGGTTCCGATACAGCAGCACCTGGTACGGTGGTGAAAGGTCCAG 967

Qy   968 GGGAGAAGCTGTTTGCTGCGCGGTGCGGCGGCCCAACTGAGCGGATGCCGCCCATCATTGCCGAAG 1027
           ||||||||||
Db   968 GTTACAAGCTTTTTGCAGCCGTTAACGCGGCCGTTAACGGAAGAACTGTGTGAGCTAAACATCATCGCAGAAG 1027

Qy   1028 ACCTGGGGGTGATCACCCCCGAGGTGGAGGCTTTGCGGCGATGGCTTCGGGTTCCCGGCA 1087
           |||||||| ||||||
Db   1028 ACCTTGGCTTCATGACGACAGTTGAAGTTGATCGAATTGATCAATGCGTAACTGAAACTGGGAA 1087

Qy   1088 TGAAGATTTTGCAGTTTGCTTTTCCGGTGAGGACAACGCCTTTTTGCCCCCACAACTACC--- 1144
           |||||||||||
Db   1088 TGAAGATTTCTTCAATTGCCTTCAACCAGGAGACGAAAGCATTGATAGCCCACACT----- 1144

Qy   1145 CCGGCACGCAATGTGGTGTGTACACGGCAAGCCACGACGACACCCTTGGGGAT 1207
           || || |||||
Db   1148 AACGTCAAAGTGGTGTGTACAGCGGCGAAGCCACGACGACACCCTTCTTGGGGAT 1207

Qy   1208 GGTTCCGCACCGCGCCGAGGCGGAGGCCTTCATGCGGGCCTACCTGGCCCGCTATG 1267
           ||||| ||||||
Db   1205 GGTACCGGAGTCGTAATGCTCGTCGTACA 1255

Qy   1268 GCATCCGTTGTTTGTCGGAATACGAGGTCGCGGGGCTTGCGCCGGCCTTCAAAA 1327
           |||||||||
Db   1256 --ACCGCTATGCTTCGTACAGTATTTTCAT 1312

Qy   1328 CGAACCCGGCCAAGCTGGCTATTGTGCCTTCGAGGACGTGTCGGGGCTGGGGCCCGAGGCCC 1387
           ||||| |||||
Db   1313 CAGTTAGCTTTATGGCAATTGCAACTATGGCAATGCAAGATTCAACTGGAAGTCTGGTTGGAAGGTGTATGGT 1372

Qy   1388 AGCAAGTGACCTTGACGCGTGCGCTACGCCGAAGGGGACC 1447
           ||||||||||
Db   1388 GCATGAACTTCCCGACGGCTGGGGGAACCGCTGGGCTACGCCGAAGGGGACC 1447

Qy   1448 TCGAGCCCGGTCTGCGCCGGGACTGCGGGCCCTGGC 1484
           |||||||||||
Db   1433 TGACACCAGCTGTCGAGGAGAAGGTTTGCTTTGACTTGAC 1469

Db   1533 TGACACCAGCTGCTGAGGAAGGTTTGCTTGACTTGAC 1569

RESULT 4
US-10-158-844-165
; Sequence 165, Application US/10158844
; Publication No. US20040029118A1
; GENERAL INFORMATION:
    APPLICANT: Kunsch et al.
    TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
    NUMBER OF SEQUENCES: 391
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: Human Genome Sciences, Inc.
        STREET: 9410 Key West Avenue
        CITY: Rockville
        STATE: Maryland
        COUNTRY: USA
        ZIP: 20850
    COMPUTER READABLE FORM:
        MEDIUM TYPE: CD-R
        COMPUTER: Dell Latitude Pentium 3
        OPERATING SYSTEM: Windows 98
        SOFTWARE: ASCII Text
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/10/158,844
        FILING DATE: 03-Jun-2002
        CLASSIFICATION: <Unknown>
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 08/961,527
        FILING DATE: 1997-10-30
        APPLICATION NUMBER: US 60/029,960
        FILING DATE: 1996-10-31
    ATTORNEY/AGENT INFORMATION:
        NAME: Hyman, Mark J.
        REGISTRATION NUMBER: 46,789
        REFERENCE/DOCKET NUMBER: PB340P1D1
INFORMATION FOR SEQ ID NO: 165:
    SEQUENCE CHARACTERISTICS:
        LENGTH: 5910 base pairs
        TYPE: nucleic acid
        STRANDEDNESS: double
        TOPOLOGY: linear
    SEQUENCE DESCRIPTION: SEQ ID NO: 165:
US-10-158-844-165

Query Match   19.1%;   Score 287.4;   DB 7;   Length 5910;
Best Local Similarity   51.7%;   Pred. No. 8.7e-70;
Matches   763;   Conservative   0;   Mismatches   696;   Indels   18;   Gaps   4;

```
Qy     11 AACGGCGTTTGAATTTTGCTCCACCCCACCAGTTTTCCGGGTCGCTGGGGATTGGGG 70
          ||||    ||    ||||||    ||||| ||||||||
Db    108 AACGTCAAAGTGGTGTGTTGATGCACATCTCTTTCTTCCAGGAGCCGCTGGAATCGGAT 167

Qy     71 CTCTGGGCCGCGGACGGCCGAGCGGGTTTTGGACTTCGTGTGATTTCTTGTCCGTGCTGT 130
          |||||||             ||||||||||    |||||||
Db    168 CATTTGGTCAAGATGCTTACGACTTCGTGATTTCTTGTTCTGTCCGTACAAAACAACGTTACT 227

Qy    131 GGCAGGTCTTACCGTGGGCCTCCACAGTTACGGCGACTCCGCCGTACCAGTCCTTCTCG 190
          ||||       ||||||||||||||||    ||||||||||
Db    228 GGCAAATCCTTCCATTAGGAGCAACTAGTTAGCGGGGATTCCTCCTTACCAATCTTCTCAG 287

Qy    191 CTTTTGCGGTAACCGTATTGGTTGACCCCGAGATGCTGATTGAAAAAGGCTGGCTGG 250
          ||||||          |||||    ||||||||||||||
Db    288 CCTTCGCAGGAAACACTCATTTTATGATTTAGTATCTTGTGGCGGTCAAGGTTTGTTGG 347

Qy    251 AACAAAGCGAAGCGCCCCCCGTATCGGCCGACCCGTGGATTGGATTATGCGGCTTTACC 310
          ||||||||||||        ||||||||||    ||||||||||
Db    348 AAGCAAGTGACCTTGAAGAGTTGGTAGCCAAGCTGTCTGAAGTTGACTATGCTA 407

Qy    311 AGACCCGCTGGCCCCCGTTGGGCGGGGGTTTCGGCGGGCAAGGCGCTTCGGCGCC 370
          |||||||             ||||||||||    |||||||||
Db    408 AAATCTACTTAGCACGTCCCTCTTTTAGAAAAAGCGGTGAAACGTTCTTCTTTGAAGTCG 467
```

```
Db    .468 GAGATGTTAAAGATTTGAGAAATTTGCTCAAGACACCAATCATGGCTTGAGCTCTTTG 527
           ||||           |||           |||||
Qy     431 CGCTCTTTATGTGCCCTCAAGACCCGTTGACGGCAAGCCGAGTGG---AGCC 487

Db     528 CTGAGTATATGGCTATCAAAGAGTATTTGCACAAATCTTGCTTGGACTGAATGGCCAGATG 587
           |||||
Qy     488 CCGAGCTGCGCGACCGTGAACCGGCTCGCCTGCCCAGGGCCCGTGAGGAGCTGGCCGAGG 547

Db     588 CAGATGCTCGTGCTCGTAAAGCTTCAGCAGTTTCTGTAACTGGTCAGCTTTGGGGTAATC 647
           ||||||||
Qy     548 AGGTGGCCCTTTACGAGCTGGATTCAGTGGCTTTTTATCTGGAATGGGCCAGACCAAGG 607

Db     648 AGTTGGTTTACCACCGTGTGACTCAAATACTTCTTCTTCCACAATGGTTGAAATTGAAAG 707
           ||||||||
Qy     608 CCTATGCCGAATCCAAGGGGATTCAGATTATGCGGATATGCCCATCTTTGTGGCCTTCG 667

Db     708 CTTACGCTAACGACAACCACACATCGAAATCGTTGGGACATGCCAATCTACGTAGCCGAAG 767
           |||||||
Qy     668 ATTCCTCAGATGTCTGGGCCAACCCGCAGTACTTCTCCGAGGCCCGATGGCAACCCCA 727

Db     768 ATTCAAGTGATATGTGGGCAAATCCAACATCCTCTTCAAAACAGATGTCAATGGTAAGGCTA 827
           |||||
Qy     728 CGGTGTGGCGCGGGCGTTCCGCGGGACTACTTCTCCCGAAACCCGCCAGCTCTGGGGCAATC 787

Db     828 CTTGTATCGCAGGATGCCCACCAGATGAGTTTTCTGTAACTGGTCAGCTTTGGGGTAATC 887
           |||||||||
Qy     788 CGCTCTATCGCTCGGGATGTGATGGAAAGGGACAACTTTGCCTGGTGCATTGCCCGCATAA 847

Db     888 CAATCTATGACTGGGAAGCAATGGACAAAGACGCTACACAAATGGTGGATTGAACGCTTGC 947
           ||||||||
Qy     848 GGCAGTGCTCAAGCAGTCGCCACCTGGTGCCGCATCGACCACTTCCCGCGGGTTTGAAGCCT 907

Db     948 GTGAAAGCTTCAAAATCTCAAAATCTACGATATCGTTCGTATCGACCACTTCCGTGGCTTCGGAATCTT 1007
           ||||||||
Qy     908 ACTGGGAGGTTCCGTTTGGCGTGGCCCAATGCGTGTGGAGGGGCGCTGGGTCAAAGCCCCAG 967

Db    1008 ACTGGGAAATCCCTGCTGTTCCGATACAGCAGCACCTGGTGAGTGGGTGAAAAGGTCCAG 1067
           ||||||||||
Qy     968 GGGAGAAGCTGTTGTCTGCTGTGCTGCGGCCCAACTGAGCGATGCGCCCATCAATGCCGAAG 1027

Db    1068 GTTACAAGCTTTTTGCAGCCGTTAAGGAAGAACTTGGTGAGCTAAACATCATCGCAGAGG 1127
           |||||
Qy    1028 ACCTGGGGGTGATCACCCCCGAGGTGGAGGCTTTCGGGTTCGCCGGCA 1087

Db    1128 ACCTTGGCTTCATGCAGGATGAAGTGATCGAATTGCGTGAAGTCGAATTCCGCTCTCAGAA 1187
           |||||
Qy    1088 TGAAGATTTGCAGTTTGCTTTTGCCTTCAACCCAGAAGACGAAAGCAGCATTGATAGCCCACACT-- 1244

Db    1188 TGAAGATTCTTCAATTTGCCTTCAACCCCCAGAAGCGGAACGCCAGCATTGATGATGATGGAT 1207
           ||||||||||
Qy    1148 CCGCGCACGCAGGCCAATGTGGTCGTACAGCGGACACGAACGCACCACCCCTGGGAT 1207

Db    1245 TGGCACCTGCTAACTCAGTTATGTGCACACAGGATACAACACGATAACAATTACGGTTCTTGTT 1304
           |||||||
Qy    1208 GGTTCCGCACCACCGGTCCGGAGCGCGGGCCCAGCGCCCTTCATGCGGGCCTACCTATG 1267

Db    1305 -----ACTCGTGAGTGAGTACATGGCTATGGCTGGCTGGCCTTCAAAA 1355
           ||||||||
Qy    1268 GGCAAATCCTTCCATTAGGAGCAACTAGTTAGCGGGGATTCCTCCTTACCAATCTTCTCAAAA 1327

Db    1356 CGAACCGTAAAGAATACGAAACAGTGGT---ACACGCTATGCTTCGTACAGTATTTTCAT 1412
           ||||
Qy    1328 GCCCGGCCAAGCTGGCTATTGTGCCTTTGCCAGGACGTGCTCAGAGACGGCAGTCTGAGGCCC 1387

Db    1413 CAGTTAGCTTTATGGCAATTGCAACTATGGCAAGATTACTAGAATTGGATGAGGCAGCTC 1472
           |||||||||
Qy    1388 GCATGAACTTCCCGACCGCTGGGGGACAACTGGGCGTTGGCGGCCTACGGCGAGGCGACC 1447

Db    1473 GTATGAACTTCCCATCTACCCTGTCGTCTTGGCGTGGAAACTGGTCTTGGCTATGCGGCTTCGGCGG 1532
           |||||||
Qy    1448 TCGAGCCCGGTTCTGGCCGCGGACTGCGGCGGGCCCTGGC 1484
```

```
Qy   371  AGGATAAGACCCGACTGGAAGCCTTCTCGAGCGCGAGCGCTTCTGGCTGGAAGACTATG  430
               ||||    ||   ||||||||||||||||||||  |||||||||||||||||
Db   468  GAGATGTTAAAGATTTGAGAAATTTGCTCAAGACACCAACCAATCATGGCTTGAGCTCTTTG  527

Qy   431  CGCTCTTTATGGCCCTCAAGACCCGGTTGACGGCAAGCCCTGGAACGCCGAGTGG---AGCC  487
               |||      ||   ||||| |||||||||||| |||||||||||||
Db   528  CTGAGTATATGGCTATCAAAGAGTATTTGCTTTGACAACTCGTTGGACTGAATGCCAGATG  587

Qy   488  CCGAGCTGCGCGACCGTGAACCGCTGCCCTTGCCCAGGGCCCGTGAGGAGCTGGCCGAGG  547
               |||   |||| |||||||||||||||||||||||||||||||  |||||||||
Db   588  CAGATGCTGCTCGTCGTAAAGCTTCAGCACTTGAAAGCTATCGTGAGCAATTGGCAGACA  647

Qy   548  AGGTGGCCCTTTACGAGTGGATTCAGTGGCTTTTTATCTGGAATGGGGCCAGACCAAGG  607
               |||||        |||||||||||||||||||||||||||||||||
Db   648  AGTTGGTTTACCACCGTGGCAATCCAATCATTGCTTCTTCCAACAATGGTGAAATTGAAAG  707

Qy   608  CCTATGCCGAATCCAAGGGGATTCAGATTATCGGCGATATGCCCATCTTTGTGGCCTTCG  667
               ||||||      ||   ||||| |||||||||||||||||||||| |||||
Db   708  CTTACGTAACGACAACCACACATCGAAATCGTTGGGACATGCCAATCTACGTAGCCGGAAG  767

Qy   668  ATTCCTCAGATGTCTCGGCCAACCCGCAGTACTTCTACCTCGAGGGCCGATGGCAACCCCA  727
               ||||      |||     ||||||||||||||||||||||||||||||
Db   768  ATTCAAGTGATATGTGGGCAAATCCACATCTTCTTCAAAACAGATGTCAATGGTAAGGCTA  827

Qy   728  CGGTGGTGGGCGGGCGTTCGCGCGGACTACTTCTTCCGGAAACCGGCCAGCTCTGGGCAATC  787
               |||   |       ||||||||||||||||||||||||||||| |||||||||
Db   828  CTTGTATCGCGACGGATGCCCACCAGATGGAGTTTTCTGTAACTGGTCAGCTTTGGGGTAATC  887

Qy   788  CGCTCTATCGCTGTGGGATGTGATGGAAAGGGACAACTTGCCTGGTGCATTGCCCGCCGCATAA  847
               |||  |        |||||||  ||||||||||||||||||||||||
Db   888  CAATCTATGACTGTGGGAAGCAATGGGACAAAGACGGCTACAAAGTTGGTGATTGAACGCGTTGC  947

Qy   848  GGCAGTCGCTCAAGCAGCAGTGCCACCTGGTCGCCATCGACCAGTTCCGCGGTTTGAAGCCT  907
               |||||||||||         |||||||||||||||||||| |||||||||
Db   948  GTGAAAGCTTCAAAATCTACGATATCGTTCGTATCGACCACTTCGTGGCTTCGTGGCTTCGAATCTT  1007

Qy   908  ACTGGGAGGTTCCGTTTGCCGGCCCGGCCCAAATGCTGTGGAGGGGCGCTGGGTCAAAGCCCCAG  967
               ||||||         |||   |||||||||||| |||||||||||
Db  1008  ACTGGGAAATCCCTGCTGGTTCCGGATACAGCAGCAACTGGTGAGTGGGTAGGGTCCAG  1067

Qy   968  GGGAGAAGCTGTGTTGCTGCGGTGCGCGTGCGGCGATGCGGCCCCATCATTGCCGCGAAG  1027
               ||||||    |||| |    ||      ||||||||||||||||||||||||
Db  1068  GTTACAAGCTTTTGCCTTGACCGGTTAAGGAAGAAACTTGTGAGCTAAACATCATCGCAGAAG  1127

Qy  1028  ACCTGGGGGTGATCACCCCCGAGGTGGAGGCTTCGCCGATGGCTTCCCGGTTCCCCGGCA  1087
               |||     |||||||||||||| ||||||||||| ||||||||||||||
Db  1128  ACCTTGGCTTCAATGACAGATGAAGTGATCGAATTGCGTGAACGTACTGGCTTCCCACGGAA  1187

Qy  1088  TGAAGATTTTTGCAGTTTTGCTTTTCCGGTGAGGACAACGCCCTTTTGTCCCCACAACTACC  1147
               ||||      ||  |||||||||||||||||||||||||   ||||||||
Db  1188  TGAAGATTCTTCAACTTTGCCTTCAACCCAGAGACGGAAAGCAGCATTGATGAGCCCACACT---  1244

Qy  1148  CCGCGCCACGGCAATGTGGTGTCGTACAGCCGGAACGGACACCAGCAACACAGACACCCCTGGGAT  1207
               |||      ||| |||||| |||||| |||||  |||||||||||||||||||||||
Db  1245  TGGCACCTGCTAACTCAGTTTATGTACACAGGAACGAATGCCAGGGCCCGTGAGGCTCGGCCGAGG  1304

Qy  1208  GGTTCCGACACCGCGCCGGAGCGCGAGCTACCTGGCCCGCTATG  1267
               ||||||         |||||||||||||||||||||||
Db  1305  GGTACCGCTAATGAGTACATGGCTCGTTACA-------ACTCGTGAGTACATGGCTCGTTACA  1355

Qy  1268  GCATCGTCGTTGTTTGTCGGAATACGAAATACAGACAGTGT---ACACGCTCATGCTTGTACAGTATTTTCAT  1327
               ||| |        ||||| ||       |     |||     ||
Db  1356  CGAACCGTAAAGCTTGTCGTAACGCTAAACAGTTACTAGAATTGGATGAAGTTCAAGTTGGATGAAG  1412

Qy  1328  GCCCCGGCCAAGCTGGCTATTTGTCCTTTGCCTTGGGGCGTGCTCGGGCTGGGGCCCCGAGGGCCC  1387
               |||||     ||  |||||||||||||||||||||||
Db  1413  CAGTTGACTTTATGGCAATTGCAACTTGGCAACTCAACTCAGAAGATTGGCGATGAGGCAGCTC  1472

Qy  1388  GCATGAACTTCCCCCGAGACGCTGGGCGTGGGGACAACTGGGCGTGGCGTACGCGCTGGACC  1447
               ||||||||| ||||       |||||| ||    ||
Db  1473  GTATGAACTTCCATCTACCCTTGGTGGAAACGTGGTCTTGGCGTATGACTGAAGATCAAT  1532

Qy  1448  TCGAGCCCCGGCCGTCTGCCGGGACTGCGCGGCCCTGC  1484
```

RESULT 5
US-10-472-928-4979/c
; Sequence 4979, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926WO
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 4979
; LENGTH: 2162598
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-472-928-4979

  Query Match   19.1%;   Score 287.4; DB 8; Length 2162598;
  Best Local Similarity 51.7%; Pred. No. 2.2e-69;
  Matches 763; Conservative 0; Mismatches 696; Indels 18; Gaps 4;

```
Qy    11  AACGGCCTTTGGAATTTGCTCACCCCACCCAGTTTCCGGTCGCTGGGGGATTGGGG  70
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1913861  AACGTCAAAGTGGTGTGTGATGCACATCTCTCTCTTCCAGGAGCTTACGGAGCCCGCTGGT  1913802

Qy    71  CTCTGGGCCGCGAGGCCGGAGCCGGTTTTGACTTCGTTGATTTCGTTGGTCCGTCGGACTTACGGGGACTTACT  130
              ||| ||||||||||||||||| ||||||||||||||||||||||||||||||
Db  1913801  CATTTGGTCAAAGTGCTTACGACTTGGCTGGGCCTACGGGGACTACCAGTCCTTCTCGG  1913742

Qy   131  GGCAGGTCTTACCGGTCGGGCCTACAGTTACGGCAACTAGTACGGGGATTCTCCTTACCAATCTTTCTCAG  190
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1913741  GGGAAATCCTTCATTAGGAGCAACTAGTACGGGGATTCTCCTTACCAATCTTTCTCAG  1913682

Qy   191  CTTTGCCGGTAACCCGTATTTGGTTGACCCGAGATGCTGATTGAAAAGGCTGGCTGG  250
              |||  ||||||||||||||||||||||||||||||||||||||||||||
Db  1913681  CCTTCGCAGGAGAAACACTCATTTTATCGATTTAGATATCTTGGTGGAGCAAGGTTTGTTGG  1913622

Qy   251  AACAAAGGCGAAGGCCCCCCGCCGTATCCGACCCAGCGCGTGAGTAGGCGATGTGTCGAGTTGGACTTACC  310
              ||| |||||||||||||||||||||||||||||||||||||||||||||
Db  1913621  AAGCAAGTGACCTTGAAGGAGTTGCGGCGGGGTTTCGGGCGCAAGGGCTTCGGGCCC  1913562

Qy   311  AGACCCGCTGGCCCCGTGTGCGCGGGGGTTTCGGGCGGGCTTCGGGCCGCC  370
              ||||||||||||||||||||||||||||||| ||||||||||||||||||
Db  1913561  AAATCTACTATGCACGTCGTCGTCTTTTTAGAAAAAGGGTGAAAACGTTTCTTTGAAGTCG  1913502

Qy   371  AGGATAAGACCCGACTGGAAGCCTTCTCGCTGCTGAAGAAGACTATG  430
              |||| ||||||||||||||||||||| |||||||||||||||
Db  1913501  GAGATGTTAAAGATTTTGAGAAATTTGCTCAAGACAACCAATCATGGCTTGAGCTCTTTG  1913442

Qy   431  CGCTCTTTATGGCCCTCAAGACCCGGTTGACGGCAAGCCCTGGAACGCCGAGTGG---AGCC  487
              ||| ||||||||||||||||| ||||||||||||||||||||||||
Db  1913441  CTGAGTATATGGCTATCAAAGAGTATTTTGACAATCTCAAAGAAGTTGAATGGCCAGATG  1913382

Qy   488  CCGAGCTGCGCGACCGTGAACCGCTGCCCTTGCCCAGGGCCCGTGAGGAGCTGGCCGAGG  547
              ||| |||||||||||||||| ||||||||||||||||||||||||||
Db  1913381  CAGATGTCGTCGTCGTAAAGCTTCAGCACTTGAAAGCTATCGTGAGCAATTGGCAGACA  1913322

Qy   548  AGGTGGCCCTTTACGAGTGGATTCAGTGGCTTTTTATCTGGAATGGAATGGGGCCAGACCAAGG  607
              |||||        |||||||| ||||||||||||||||||||||||
Db  1913321  AGTTGGTTTACCACCGTGGCTCAATCAATACTTCTTCCAACAATGGTGAAATTGAAAG  1913262

Qy   608  CCTATGCCGAATCCAAGGGGATTCAGATTATCGGCGATATCGGCCATCTTTGTGTGGCCTTCG  667
              |||  |       |||||||||||||||||||||||||||||||| ||||||||
Db  1913261  CTTACGTTAACGACAACCACCGTGTTCGGTTCTTGGCGAAATCGAAAATCGTGGGGACATGCCAATCTACGTAGCCGGAAG  1913202

Qy   608  CCTATGCGAATCCAAGGGGATTCAGATTATCGGCGATATCGGCCATCTTTGTGTGGCCTTCG  667
```

```
ZIP: 02354
COMPUTER READABLE FORM:
   MEDIUM TYPE: CD/ROM ISO9660
   COMPUTER: <Unknown>
   OPERATING SYSTEM: <Unknown>
   SOFTWARE: <Unknown>
CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/10/617,320
   FILING DATE: 10-Jul-2003
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: US/09/107,433
   FILING DATE: 30-Jun-1998
   APPLICATION NUMBER: 60/ 085131
   FILING DATE: May 12, 1998
   APPLICATION NUMBER: 60/051553
   FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
   NAME: Ariniello, Pamela Deneke
   REGISTRATION NUMBER: 40,489
   REFERENCE/DOCKET NUMBER: GTC-011
TELECOMMUNICATION INFORMATION:
   TELEPHONE: (781)893-5007
   TELEFAX: (781) 893-8277
INFORMATION FOR SEQ ID NO: 336:
   SEQUENCE CHARACTERISTICS:
      LENGTH: 1377 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: double
      TOPOLOGY: circular
   MOLECULE TYPE: DNA (genomic)
   HYPOTHETICAL: NO
   ANTI-SENSE: NO
   ORIGINAL SOURCE:
      ORGANISM: Streptococcus pneumoniae
   FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (B) LOCATION 1...1377
   SEQUENCE DESCRIPTION: SEQ ID NO: 336:
US-10-617-320-336
```

```
Query Match        17.7%;   Score 265.6;  DB 9;  Length 1377;
Best Local Similarity   52.1%;   Pred. No. 8.5e-64;
Matches  698;  Conservative   0;  Mismatches 624;  Indels  18;  Gaps   4;
```

```
Qy  148  GGCCCTACCAGTTACGGCGACTCGCCGTACCAGTCCTTCTCGCGTTTGCCGGTAACCCG  207
             |||||     ||||||||||||||||||||| ||||||| || |||| |
Db    4  GGAGCAACTAGTTACGGGGATTCTCCTTACCAATCTTCTCAGCCTTCGCAGAAACACT   63

Qy  208  TATTTGGTTGACCCCGAGATGCTGATTGAAAAAGGCTGCTGGAACAAAGCGAAGCGCCC  267
             |||||| | |||     || ||| ||   ||||     ||||| ||
Db   64  CATTTTATCGATTTAGATATCTTGGTGGAGCAAGGTTTGTTGGAGCAAGCAAGTGACCTTGAA  123

Qy  268  CCGCGTATCCGACCCAGCGCGTGGATTATGCGTGGATTATGCTTGCGTTACCAGACCCCGTGCCCCTG  327
             ||||         |||      |||| ||| ||||  || |||||||||
Db  124  GGAGTTGACTTTGGTAGCGATGCGTCGAGTTGACATGCTCTAAAATCTACTATCGACTG  183

Qy  328  TTGCGGCGGGTTTCGGGGTTTGAGAAACGCGGTGAAACGTTTCTTTGCTCGAAGACTATGCCTCTTTATGGCCCTC  387
Db  184  CGTCCTCTTTTTGAGAAAAAGCGGTGAAACGTTTCTTTGAAGTCGGAGATGTTAAAGATTTT  243

Qy  388  GAAGCCTTTATCGAGCGCAGTCGTGCTCGGGTCCGGAGGCCC  447
Db  244  GAGAAATTTGCTCAAGACAACCAATCATGGCTTGAGTAGTATGCTATC  303

Qy  448  AAGACCCGGTTTGACGGCAAGCCCTGGAACGCAGTGGAGC---CCCGAGTGCGCGACCGT  504
Db  304  AAAGAGCATTTTGACAATCTTGCTTGGACTGGACCAGATGGCCAGAATGCAGATGCTCGTCGT  363

Qy  505  GAACCGGCTGCCCTGGCCAGGGCCCGTGGAGAGCTGGCCGGAGGAGGTGCGCCCTTTACAG  564
Db  364  AAAGCTTCAGCACTTGAAAGCTATCGTGAGCAATTGGCAGACAAGTTGGTTTACCACCGT  423

Qy  565  TGGATTCAGTTAGTCGGTCTTTTTTATCTGAATGGGGCCCTATGCCGAATCCAAG  624
```

```
Qy  668  ATTCCTCAGATGCTCTGGGCCAACCCGCAGTACTTCTACCTCGAGGCCGATGGCAACCCA  727
             |||||
Db  1913201 ATTCAAGTGATATGTGGGCAAATCCACATCTCTTCAAAACAGATGTCAATGGTAAGGCTA  1913142

Qy  728  CGGTGGTGGCGGGCGTTCGCGGGACTACTTCTCCGAAACCGGCCAGCTCTGGGGCAATC  787
Db  1913141 CTTGTATCGCAGGATGCGCCACCAGATGAGTTTCTGTAACTGTGTCAGCTTTGGGGTAATC  1913082

Qy  788  CGCTCTATCGCTGGGATGTGATGGAAAGGGACAAACTTTGCCTGGTGCATTGCCCGCATAA  847
Db  1913081 CAATCTATGACTCGGAAGCAATGCAAAGACGGCTACAAAATGGTGGATTGAACGCTTGC  1913022

Qy  848  GGCAGTCGCTCAAGCAGTGCCACCTGGTGCGCATCGACCACTTCCGCGGGTTTGAAGGCT  907
Db  1913021 GTGAAAGCTTCAAAATCCAAAATCACGTTCGTATCGACCACTTCCGTGGCTTCGAATCTT  1912962

Qy  908  ACTGGGGAGGTTCCGTTTGCGCCCAATGCTGTGGAGGGGCGCTGGGTCAAAGCCCCAG  967
Db  1912961 ACTGGGAAATCCCTGCTGGTTCCGATACAGCAGCAGCTGGTGAGTGGGTGAAGGTCCAG  1912902

Qy  968  GGGAGAAGCTGTTTGCTGCGGGTGCGGGCCCAACTGAGCGATGCGCCATCATTGCCGAAG  1027
Db  1912901 GTTACAAGCTTTTTGCAGCCGTTAAGGAAGAACTTGGTGAGCTAAACATCATCGCAGAAG  1912842

Qy  1028  ACCTGGGGGTGATCACCCCGAGGTGGAGCGCTTGCGCGATGGCTTCCCCGGCA  1087
Db  1912841 ACCTTGGCTTCATGACAGATGAAGTGATCGAATTGCGTGAACGTACTGGCTTCCCAGGAA  1912782

Qy  1088  TGAAGATTTGCAGTTTGCTTTTCCGGTGAGGACAACGCCTTTTGCCCCACAACTACC  1147
Db  1912781 TGAAGATTCTTCAATTTGCCTTCAACCCAGAAGAGAAAGCATTGATAGCCCACACT---  1912725

Qy  1148  CCGCGCACGGCAATGTGTGGTGTACAGCGGAACCCACGACAAACGACACCACCCCTGGGAT  1207
Db  1912724 TGGCACCTGCTAACTCAGTTATGTACACAGGGACACACGATAACAATACGGTTCTTGGTT  1912665

Qy  1208  GGTTCCGCACCGCGCCGAGGCGGGCCTTCATGCGGGCCGCCTATG  1267
Db  1912664 GGTACCGTAATGAGATTGATGATGCG--------ACTCGTGAGTACACAGGGCTCGTTACA  1912614

Qy  1268  GCATCCGTTGTGTTTGTCGGAAATACGAGAGGTTCGCGGGCGCTTGATCGAGCTGGCCTTCAAAA  1327
Db  1912613 CGAACCGGTAAAGATAACGAAACAGTGGT---ACACGCTATGCTTCGTACAGTATTTTCAT  1912557

Qy  1328  GCCCGGCCAAGCTGGCTATTGTGCCTTTGCAGGACGTGCTCGGGGCTGGCCCCCGAGGCCC  1387
Db  1912556 CAGTTAGCTTTATGGCAATTGCAACTATGGCAACTATGGATTTACTAGAATTGGATGAGGCAGCTC  1912497

Qy  1388  GCATGAACTTCCCCGGACGCGGTCGGGGGACCAACTGGCGGTCGCGCCGCTACGCGGCGACC  1447
Db  1912496 GTATGAACTTCCCATCTACCCTTGGTGGAAACTGGCTGGTCTTGGCGCTATGACTGAAGATCAAT  1912437

Qy  1448  TCGAGCCCGGTCTGGCCGCGGGACTGCGCGGGCCCTGGC  1484
Db  1912436 TGACACCAGCTGCGCTGGCAGAGGTTGCTTGCTACTTGAC  1912400
```

```
RESULT 6
US-10-617-320-336
; Sequence 336, Application US/10617320
; Publication No. US20050136404A1
; GENERAL INFORMATION:
;   APPLICANT: Lynn A Doucette-Stamm and David Bush
;   TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
;                      SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  FOR DIAGNO
;                      THERAPEUTICS
;   NUMBER OF SEQUENCES: 5206
;   CORRESPONDENCE ADDRESS:
;      ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;      STREET: 100 Beaver Street
;      CITY: Waltham
;      STATE: Massachusetts
;      COUNTRY: USA
```

; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 31307
; LENGTH: 1676
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73152G04_FLI
US-10-425-114-31307

Query Match   10.6%;   Score 159;   DB 7;   Length 1676;
Best Local Similarity 50.5%;   Pred. No. 5.4e-34;
Matches 413;   Conservative 0;   Mismatches 402;   Indels 3;   Gaps 1;

Qy   415 TGGCTCGAAGACTATGGCGTCTTTATGGCCCTCAAGACACCCGGTTGACGGCAAGCCCTGG 474
                ||||| ||| || ||||| ||||||||| ||| ||| |||||||||| ||||||||
Db   379 TGGCTTGAAGATGCAGCTCTTTTTTGCTCATTGACAAAGCATCAATGACATCTCTTGG 438

Qy   475 AACGAGTGGAGCCCCAGTCTGCGACCGTGAACCGGCTGCCCTGGCCAGGGCCCGTGAG 534
                |||||| ||| || |||| ||| |||||| || ||| |||||||
Db   439 TACGAGTGGCCTGAACCTTGGAAAAATCGCCATCTTGGAAGATATATATGAA 498

Qy   535 GAGCTGGCCGAGGAGTGGCCCTTTACGAGTGGATTCAGTGGCTTTTTATCTGGAATGG 594
                |||| || |||||| ||||| ||||| || || |||| |||||||||
Db   499 AAACAGAAGGATTTTATAGACATTTATGGCTCAGCAGTTCTTATTTCAAAGCAATGG 558

Qy   595 GGCCAGACCAAGGCCATGCGAATCCAAGGGATTCAGATTATCGCGATATGCCCATC 654
                || ||||||| ||| || |||||| ||||||||||| |||| ||||||
Db   559 CAGCGGATTCGTAAATATGCACAAAGCTGGTATCAGCATCATGGGTGACATGCCTATA 618

Qy   655 TTTGTGCCTTCGATTCCTCAGATGTCTCGGCGGCGGTTCCGCGGACTACTTCTACCTCGAGGCC 714
                |||| ||||||| |||||| |||||| | ||||||| ||||| |||| ||||||
Db   619 TATGTTGGCTACCATAGCGCCAGATGTTTGGCGGCGAACAAGGAAATCATTTTTGCTGGACAAA 678

Qy   715 GATGGCAACCCACGGTGGTCGGTTCCGCGGGACTGGAGTTCCGACGGCGCCG 774
                || |||||||| |||| || |||| || |||| || || ||||||
Db   679 AACGGTTTCCCCACTTTCGTTAGTGCAGTTCCACCTGATGCATTTAGTGAAACGGGTCAA 738

Qy   775 CTCTGGGGGCAATCGCTCTATCGCCTATGGCATCCGTTGTTTGTGC 834
                ||||| ||| ||||| || ||||| |||| |||||||||||
Db   739 CTATGGAAGCAGTCCATTGTACGACTGGAAAGCTATGGAAGCTATGGTGTTCATGGTGG 798

Qy   835 ATTGCCCGCATAAGGCAGTGGCCTCAAGCAGTGCCACCTGGTGCGCATCGACCACTTCCGC 894
                |||||| |||||| || |||| |||| || ||||| |||||||| |||||||
Db   799 ATAAAGAGGATTAAACGTGCCCTTGATTTGTATGATGAAGTTCCGATTGACCATTTCCGG 858

Qy   895 GGGTTGAAGCCTACTGGGAGGTTCCGTTTGGCCGGCCCAATGCTGTGGAGGGGCGCTGG 954
                ||| |||||| ||| ||| ||||| |||| |||||||| || |||||| ||||
Db   859 GGGCTTGCGGGTTTTTTGGGCAGTCCCTTCTGAAAGCTGGTCAAAAGTAGCACTGGTTGGAAGCTGG 918

Qy   955 GTCAAAGCCTGGACCGGCAGAGAGAGCTGTTGCTGCGTCGCCACCTGGTCGCATCGAGCGATGCGCCC 1014
                ||||||| ||| |||| || |||||||||||||| |||||||||||||||||||
Db   919 AGGGCTGGACCAAGAAGAATAGCTTTTTGACACGCTCTTCAAAGCTGGTCGTTGGTAGAATAGAT 978

Qy   1015 ATCATTGCCGAAGACCTGGGGTGATCACCCCGAGGTGGAGGCTTTTGCGCGATGGCTTC 1074
                ||||||| |||||||| ||| |||| ||||||||||||||| |||| ||||||||
Db   979 ATAATAGCAGAGATCTGGGGGTAATTACTGACAGTCGTTCAGCTAAGAAATCCATT 1038

Qy   1075 GGGTTCCCCGGCATGAAGATTTGCAGTTGCTTTGTTTCCGGTGAGGACAACGCCTTTTTG 1134
                ||||||| ||| |||| |||| || || |||||| || || ||||||||||
Db   1039 GGGGCCCCTGGGATGGCAGTTCTCCAGTTTGTTTCGGGTGGTTCTGCACACCCTCAT 1098

Qy   1135 CCCCACAACTACCCCGCACCGCAATGTGGTGGTTGTACAGCGGAAT 1194
                |||||| || |||| || |||||| | || ||||||| |||
Db   1099 TTGCCACACACCATG---AAATGGATCAAGTTGTGTACACTGGAACACATGATAACGAT 1155

Qy   1195 ACCACCCTGGGATGGTTCCGCACCACCTGGTCCGGAGGCCGA 1232

Db   1156 ACAGTTCTTGGCTGGTGGCAAAATTTACCAGAGGAAGA 1193

RESULT 8
US-10-425-115-140665
; Sequence 140665, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;   APPLICANT: La Rosa, Thomas J.
;   APPLICANT: Kovalic, David K.
;   APPLICANT: Zhou, Yihua
;   APPLICANT: Cao, Yongwei
;   TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;   TITLE OF INVENTION: Plants
;   FILE REFERENCE: 38-21(53222)B
;   CURRENT APPLICATION NUMBER: US/10/425,115
;   CURRENT FILING DATE: 2003-04-28
;   NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 140665
;   LENGTH: 1788
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_59770C.1
US-10-425-115-140665

Query Match   8.9%;   Score 134; DB 8; Length 1788;
Best Local Similarity 52.4%;   Pred. No. 5.3e-27;
Matches 319; Conservative   0; Mismatches 287; Indels   3; Gaps   1;

Qy   624 GGGGATTCAGATTATCGGCGATAATGCCATCTTGTGCCTTCGATTCCTCAGATGTCTG 683
               ||||||| ||| || |||||| ||| | || |||| | |||||| |||||
Db   586 GCGTATCAGCATCATCGGTGACATGCCTCTATATGCTGGCTACCATAGCGCATATGTTTG 645

Qy   684 GGCCAACCCGCAGTACTTCTACCTCGAGGCGCCGATGGCAACCCCACGGTGGTGGCGGCGT 743
               ||||||| || ||| | |||| |||| |||| |||||| | ||||||| ||||
Db   646 GGCGCAACAGGAAATCATTTTTGCTGGACAAAAACGGTTTCCCCACTTTCGTTAGTGCGGT 705

Qy   744 TCCGCGGGACTACTTCTCCGAAACCGGCCAGCTCTCGGGCAATCCGCTCTATCGCTGGGA 803
               || ||| || ||| | || |||||||| | | ||||| | ||||| |||||||
Db   706 TCCACCTGATGCATTTAGTGAAACGGTCAACTATGAACAGTCCATTGTACGACTGGAA 765

Qy   804 TGTGATGGAAAGGACAACTTTGCCTGGTGCATTGCCCGGCATAAGGCAGTCGCTCAAGCA 863
               ||| ||||||| ||||||| | || | || ||||||| |||| | | |||||
Db   766 AGCTATGCAAGGACAGATTGTTTTTCATGGTGGATAAAGAGGATTAAACGTGCCCTTGATTT 825

Qy   864 GTGCCACCTGGTGCGCATCGACCACTTCCGCGGGTTGAAGCCTACTGGGAGGTTCCGTT 923
               ||||||||| || | | |||||| | || ||||| ||||| | | |||||
Db   826 GTATGATGAATTCCGTATTGACCATTTCGCGGGGTTTTCGGGCCAGTCCCTTC 885

Qy   924 TGGCCGCCCAATGCTGTGGAGGGGGCCGCTGGTCAAAGCCCCAGGAGGAGAAGCTGTTTGC 983
               |||||||||| | ||||||||||||| || ||||| ||||||||||||| || |
Db   886 TGATGCAAAAGTAGCACTGGTTGGAAGCTGAGGGCTGAGGGAAGGAATAGCTTTTTTGA 945

Qy   984 TGCGGTGCGGGCCAACTGAGCGATGCGCCCATCATTGCCGAGACCTGGGGTGATCAC 1043
               ||||||||||||| | |||||| | || ||||||||||| |||||||||||||
Db   946 CACGCGTCTTCAAAGCTGTTGGTAGAATAGATAATAATAGCAGAAGATCTGGGGTAATTAC 1005

Qy   1044 CCCCGAGGTGGAGGGTTCGTTCAGCTAAGGAAATCCATTGGGGCCCCTGGGATGGCAGTT 1103
               ||||||| | |||||| | ||| | ||||| | ||||| |||| |||
Db   1006 TGAAGATGTCGTTCAGCTAAGGAAATCCATTGGGGGCCCCTGGGATGGCAGTTCTCCAGTT 1065

Qy   1104 TGCTTTTTCCGGTGAGGACAACGCCCTTTTTGCCCCAACTACCCCCGCGCACGGCAATGT 1163
               ||||||||||||| ||||| ||||| | ||||| || || | | | | ||
Db   1066 TGCTTTTTCCGGAGGTGGTTCTGACAAACCCTCATTTGCCCACACACCATG--AAATGGATCA 1122

Qy   1164 GGTGGTGTACAGCGGAACCACGACACCCTGGGATGGTTCCGCACCGCGCC 1223
               || |||| || | | |||||| ||| |||||||| | |||||||
Db   1123 AGTTGTGTACACTGGAACACACTGATAACGATAACGATACAGTTCTTGGCTGGTGGCAAAAATTTACC 1182

Qy   1224 GGAGGCCGA 1232
               ||| ||

; OTHER INFORMATION: Clone ID: PAT_MRT3847_125408C.1
US-10-424-599-28136

Query Match   6.8%;   Score 99.2; DB 7;   Length 2540;
Best Local Similarity   47.6%;   Pred. No. 3e-17;
Matches 370; Conservative   0; Mismatches 393; Indels 15; Gaps   2;

Qy   22 GGAATTTGCTCCACCCACCAGTTTCCGGGTCGCTGGGGGATTGGGGCTCTCGGGCCGC   81
Db   332 GGCATTCTGCTCCACCCGAGTCGTTTCGCGGGCCCCTACGGAATTGGCGATCTTGGCGAC 391

Qy   82 GAGGCCGAGCCGGTTTTTGGACTGGCTGGCGTCCACCACTCGGGTTGTCTCCGTTTGGCAGGTTCTT 141
Db   392 GAGGCGGTTCCGTTCATCGACTGGCTCCACCACTCGGGTTGTCTCCGTTTGGCAGGTTCTT 451

Qy   142 CCGCTGGGCCCTA--------CCAGTTACGGCGACTCGCCGTACCAGTCCTTCTCG 189
Db   452 CCCCTCGTGCCACCTGGCAGAAAGGCCAACGAAGGATCCCCTTATTCTGCCAGGAT 511

Qy   190 GCTTTTGCCGGTAACCCGTATTTGGTTGACCCCGAGATGCTGATTGAAAAAGGCTGGCTG 249
Db   512 GCGAATTGTGGCAACACGCTCTTGATCTCTTTGAGGGCCTCGTCGAGGATGGGTTGTTG 571

Qy   250 GAACAAAGCGAAGCGGCCCCCGCCGTATCGCACCTGGTCGCGCATCG 309
Db   572 GAAAAACACGAGCTTCCTCAACCAATTGATGCAGAGCGCGTGAATTTTTCACTTGTTGCT 631

Qy   310 CAGACCCGCTGGCCCCTGTTGCCGGAGGTTCCGTTTGGCCGGGCCCAATGCTGTGTGG 369
Db   632 GATGTTAAGGATCCTTTGATAACTAAGGCTCCAGAGAGCCTCATTTCAAGTGAAGGGGAA 691

Qy   370 CAGGATAAGACCCGACTGGAAGCCTTTATCGAGGCCGAGCGGCT---TCTGGCTGGAAGAC 426
Db   692 CTCAAAACACAGCTTGAGAATTTTCGTCGTGACCCTGACATATCAAGCGTGCTTGAAGAT 751

Qy   427 TATGCGCTCTTTATGCGCCTCAAGACCCGGTTGACGGCAAGCCTGGCCCGGAGGCTTTGC 486
Db   752 GCAGCCTATTTTGCTCGTATTGACGACAGTTTAAACAACAATCAGCTGTACAATTGGCCT 811

Qy   487 CCCGAGCTGCGCGACCTGAACCGGCTGCCCTGGCCAGGGCCCGTGAGGAGCTGCCGAG 546
Db   812 GAACCTTAAGAAATCGCCATCTTGTAGCTCAGAAGATATTATTCACCAACAAAGCGAGAT 871

Qy   547 GAGGTGGCCCTTTACGAGTGGATTCAGTGGCTTTTTATCTGAATGGGGCCAGACCAAG 606
Db   872 TTTATAAATGTATTTATTGCCCAACAGTTCTTGTTTCAAAGGCAGTGGCAGAAAGTTCAT 931

Qy   607 GCCTATGCGAATCCAAGGGATTCAGATTATCGGCGATATGCCCATCTTTGTGGCCTTC 666
Db   932 AGCTATGCACAGAGTAAGGAATCAGCATAATGGGAGACATGCCAATATATGTTGGTTAT 991

Qy   667 GATTCCTCAGATGTCTGGGCCAACCCCAGTACTTCTACCTGGCAATAAGAAAACAGTTTTT 726
Db   992 CATAGCGCAGATGTTGGCAATAAGAACAAGTTTTTACTGACGAGAAACCGGCCAGCTCTGGGGCA 1051

Qy   727 ACGGTGGTGCGGGGCGTTCCGCGGGACTACTTCTCCGAAACCGGCCAGCTCTGGGCA 784
Db 1052 CTTTTAGTTAGTGGTGTTCCTCTGACGCATTCAGTGAAACCGGCCAGCTCTGGGGCA 1109

RESULT 12
US-10-156-761-5375
; Sequence 5375, Application US/10156761
; Publication No. US2003011901A1
GENERAL INFORMATION:
..; APPLICANT: OMURA, SATOSHI
..; APPLICANT: IKEDA, HARUO
..; APPLICANT: ISHIKAWA, JUN
..; APPLICANT: HORIKAWA, HIROSHI
..; APPLICANT: SHIBA, TADAYOSHI
..; APPLICANT: SAKAKI, YOSHIYUKI
..; APPLICANT: HATTORI, MASAHIRA
..; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

..; APPLICANT: Barbazuk, Brad
..; APPLICANT: Li, Ping
..; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
..; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
..; FILE REFERENCE: 38-21(53221)B
..; CURRENT APPLICATION NUMBER: US/10/437,963
..; CURRENT FILING DATE: 2003-05-14
..; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 36082
; LENGTH: 886
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_39941C.1
US-10-437-963-36082

Query Match   6.8%;   Score 102.2; DB 7;   Length 886;
Best Local Similarity   52.6%;   Pred. No. 3.7e-18;
Matches 247; Conservative   0; Mismatches 220; Indels 3; Gaps   1;

Qy   764 AAACCGGCCAGTCTCGGGCAATCCGCTCTATCGCTCTGGGATCTGGGATCGTGATGATGTGATGGAAAGGGACAACT 823
Db   5 AAACTGGTCAGCTTTGGAACAGTCCATTGTACGACTTGGAAAGCTATGGAAGCAGTGGAT   64

Qy   824 TTGCCTGGTGCATTGCCCGCATAAGGCAGTCGCTCAAGCAGTGCCACCTGGTGCGCATCG 883
Db   65 TTGAATGGTGGATAAAAAAGGATTAATCGTGCCCTTGATTGTGATGATGAGTTCCGTATAG 124

Qy   884 ACCACTTCCGCGGGGTTGAAGCCTACTGGAGGTTCCGTTTGGCCGGGCCCAATGCTGTGG 943
Db   125 ACCATTTCCGGGGGCTTGCAGTTTTTGGGCAGTTCCTTCTGAGTCAAAAGTAGCGCTGG 184

Qy   944 AGGGGGCGTGGGTCAAAGCCCAGGGAGGAGAAGCTGTTGCTGCCGTGTCGCGGGCCCAACTGA 1003
Db   185 TTGGAAGCTGGAGGGGCTGGACCGGAGGAATGCATTTTTGATGCATTATTCAAAGTGTCG 244

Qy 1004 GCGATGCGCCCATCATTGCCGGAAGACCTGGGGTGATCACCCCCGAGGTGGAGGCTTTGC 1063
Db   245 GTAGAATAAACATAATAGCAGGAAGACCTGGGGTGATAACGAAGATGTTGTTGACCTAA 304

Qy 1064 GCGATGGCTTCGGGGTTCCCCGGCATGAAGATTTGCAGTTGCTTTTCCGGTGAGGACA 1123
Db   305 GGAAGTCTATTGAGGCTCCTGGAATGCAGTTCTTCAGTTTGCTTTTGTCTTTTTGGCGTGGTTCTG 364

Qy 1124 ACGCCTTTTGCCCCACAACTACCCCGCGCAATGTGTGGTGTACAGCGGAACCC 1183
Db   365 ACAATCCACACTTGCCCCATAACCATG---AATTTGATCAAGTAGTATACACCGGAACAC 421

Qy 1184 ACGACAACGACACCACCTGGATGGTTCCGCACCGGCGCCGAGGGCCGAG 1233
Db   422 ACGACAATGACACGGTTATTGGTTGGTGGCAAACTTTACCAGAGGAAGAG 471

RESULT 11
US-10-424-599-28136
; Sequence 28136, Application US/10424599
; Publication No. US20040031072A1
GENERAL INFORMATION:
..; APPLICANT: La Rosa Thomas J
..; APPLICANT: Kovalic David K
..; APPLICANT: Zhou Yihua
..; APPLICANT: Cao Yongwei
..; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
..; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
..; FILE REFERENCE: 38-21(53223)B
..; CURRENT APPLICATION NUMBER: US/10/424,599
..; CURRENT FILING DATE: 2003-04-28
..; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 28136
; LENGTH: 2540
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:

Qy 1168 GTGTACAGCGGAACCCAGACGACAACGACACCACCCTGGGATGTTCCGCACCGCGC 1222
         ||| |||||| |||
Db 1732 GCCACCGCGCCACCACCCCACGACCTGCCGCCGCCGACCGCCGCCGCCTCACCGCGGCCGC 1786

RESULT 13
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match   5.6%;   Score 84.2;   DB 6;   Length 9025608;
Best Local Similarity   46.6%;   Pred. No. 1.7e-12;
Matches   389;   Conservative   0;   Mismatches 428;   Indels 18;   Gaps   3;

Qy   391 GCCTTTATCGAGGCCGAGCGCTTCTGGCTGGAAGACTATGCGGCTCTTTATGCCCTCAAG 450
Db 6537138 GACTTCCTCGCCGAGGAGGGCGAGGCGCTGGAGGACCACCGCACGTGGTGCGGCGCT---G 6537194

Qy   451 ACCCGGTTTGACGGCAAGCCCTGGAGCTGGAGCACGGTCTGCGCGGCTGCCGCCGTGAACCG 510
Db 6537195 GCCGAGGTGTACGGGCCCGATGGGCACGGGCTGGCCGGAGGTGGCCCTTTACGAGTGATT 6537254

Qy   511 GCTGCCCTGGCCAGGGCCCGTGAGGAGCTGGCCCTTTACGAGTGGATT 570
Db 6537255 GCCGAGACCGCCGGGGCGGCGGAGTTGATGGATCGACCAAGGCTATGCCGCCGGCTC 6537314

Qy   571 CAGTGGCTTTTTATCTGGAATGGGCCAGACCAAGGCCTATGCCGCCAAGGGGATT 630
Db 6537315 GCCTGGGCTGACCGACGAGCAGCTCGCCGCCGGCGTCAGCGAGCGTGTGGGCCATG 6537374

Qy   631 CAGATTATCGGGCGATATGCCCATCTTTGTGGCCTTCGATTCCTCAGATGTCTGGGCCAAC 690
Db 6537375 GCGGTCGGACTCGTGCACGACCTCGCGGTCGGCGTCATCCCGTCGGCGGACGCCTCG 6537434

Qy   691 CCGCAGTACTTCTACCTCGAGGCCGATGGCAACCCCACGGTGGTGGGCGGGCGTTCCGCGG 750
Db 6537435 GCGCAGCAGGAGTACTTCGCCGCGGGCA--------TGTCGGTGGGCGGCCCCGCCCC 6537482

Qy   751 GACTACTTCTCCGAAACCGGCCAGCTCTTGGGGCAATCCGCTCTGGGATGTGATG 810
Db 6537483 GACGCCTTCAACGCGCTCGGCCAGGACTGGGCGGCCTGCCCCGGGACCGGACCGCCTC 6537542

Qy   811 GAAAGGGACAACTTTGCCCGTGATTGCCCGCCATAAGCAGTCGCTCAAGCAGTGCCAC 870
Db 6537543 GCCGAGTCCGGCTACGCGCCCGTACCCGGCCTGCTCCCGCGTCTGCGATACGCGGGC 6537602

Qy   871 CTGGTGCGGCATCGACCACTTCCCGCGGGTTTGAAGCCTACTGGGAGGTTCCGTTTGGCCGG 930

```
Db   6537603  GCCCTGCGCATCGACCACGTGATGGGGCGCTGTTCCGGCTGTGTGGTGGTGCCCAGGGGCGG  6537662
                |||||||||||||||  |||  |||   |||  ||| ||
Qy      931    CCCAATGTCGTGGAGGGGCGCTGGTCAAAGCCCAGGGGAGAAGCT---GTTTGCTGCG      987

Db   6537663  CCGCCCACGGAGGGAACGTACGTCCGGTACGACGGCCCAGGGCGATGCTTCGCGATCCTCGTC  6537722
                ||||||||  |                |||||
Qy      988    GTGCGGGCCAACTGAGCGATGCGCCCATCATTGCCGAAGACCTGGGGGTGATCACCCCC     1047

Db   6537723  CTGGAGGCATCGCGCACCGGGGCACTGTGATCGGCGAGGACCTCCGCACGGTCGAACCC  6537782
                ||||  ||    |  ||   |||
Qy     1048    GAGGTGGAGGCTTGCGCGATGGCTTCGGGTTCCCCGGCATGAAGATTTGCAGTTTGCT     1107

Db   6537783  GGGGTGCGCGAGAGCGTGCACGCGCGGGCGTGCTCGGCACGTCGGTGCGTGGTTCGAA  6537842
                  ||  |  | |||
Qy     1108    TTTTCCGGTGAGGACAACGCCCTTTTGCCCCAACTACCCCGCGCACGGCAATGTGGTG     1167

Db   6537843  CGGGACTGGGAGGGGGACGGGCGTGCCCGTGCCGCCCGCGCTGGCCGCGGACTGCCTG  6537902
                |||    ||
Qy     1168    GTGTACAGCGGAACCCAGCACGACAACGACACCACCCTGGATGGTTCCGCACCGCGC     1222

Db   6537903  GCCACCGCACCCACCACCCCGACCTGCCCGCGCGACCGCCCGCCGCCTCACCGGGCCGC  6537957
```

```
Query Match             4.4%;   Score 66.8;  DB 7;  Length 1945;
Best Local Similarity  57.0%;   Pred. No. 3.2e-08;
Matches  122;  Conservative   0;  Mismatches  92;  Indels   0;  Gaps   0;

Qy    571   CAGTGGCTTTTTATCTGGAATGGGGCCAGACCAAGGCCTATGCCGAATCCAAGGGGATT   630
             ||||||   ||||||  ||||  ||||||  |||    |||||
Db    326   CAGTTCCTGTTTCAAAGGCAGTGGCAGAAAGTTCATAGCTATGCACAGAGTAAGGGAATC  385

Qy    631   CAGATTATCGGCGATATGCCCATCTTTGTGGCCTTCAGATGTCTCGGCCAAC           690
             ||||    |  || ||   ||||||||  ||||  |||
Db    386   AGCATAATGGGAGACATGCCAATATATGTTGGTTATCATAGCGCAGATGTTTGGCAAAT   445

Qy    691   CCGCAGTACTTCTACCTCGAGGCCGATGGCAACCCCACGGTGGCGGGCGTTCCGCGG   750
             ||||   |  |  ||    |
Db    446   AAGAAACAGTTTTTACTGAACAGGAAGGGCTTTCCTCTTTAGTTAGTGGTGTTCCTCCT  505

Qy    751   GACTACTTCTCCGAAACCGGCCAGCTCTGGGCA                           784
             ||||   |||||||||  ||||  |||||
Db    506   GACGCATTCAGTGAAACCGGCCAGCTCTGGGCA                          539
```

Search completed: January 14, 2006, 17:18:12
Job time : 1164 secs

```
Query Match             5.1%;   Score 76.8;  DB 7;  Length 563;
Best Local Similarity  53.8%;   Pred. No. 4.3e-11;
Matches  156;  Conservative   0;  Mismatches 134;  Indels   0;  Gaps   0;

Qy    571   CAGTGGCTTTTTATCTGGAATGGGGCCAGACCAAGGCCTATGCCGAATCCAAGGGGATT   630
             ||||    ||||||  ||||  |||||   |||
Db    252   CAGTTCCTGTTTCAAAGGCAGTGGCAGAAAGTTCACAGCTATGCACAGAGTAAGGGAATC  311

Qy    631   CAGATTATCGGCGATATGCCCATCTTTGTGGCCTTCAGATGTCTCGGCCAAC           690
             ||||    | ||| ||   ||||||||  ||||  |||
Db    312   AGCATAATGGGAGACATGCCAATATATGTTGGTTATCATAGCGCAGATGTTTGGCAAAT   371

Qy    691   CCGCAGTACTTCTACCTCGAGGCCGATGGCAACCCCACGGTGGCGGGCGTTCCGCGG   750
             |||    | |  ||  |
Db    372   AAGAAACAGTTTTTACTGAACAGGAAGGGCTTTCCTCTTTAGTTAGTGGTGTTCCTCCT  431

Qy    751   GACTACTTCTCCGAAACCGGCCAGCTCTGGGCAATCCGCTTATCCGCTGGGATGTGATG   810
             ||||   |||||||||  ||||  |||||
Db    432   GACGCATTCAGTGAAACCGGTCAGCTTTGGGCCAGCCCCTGTATCATGATTGATTG... GAAAGCCATG  491

Qy    811   GAAAGGGACAACTTTGCCTGCGTGCATTGCCCGACGCATAAGGCAGTCGCTCAA         860
             ||||   ||
```

```
Db    492   GAGAAAGATGGATACCTCATGGGGTACGTNGCATACGACGNGCCACAAA   541
```

```
                    GenCore version 5.1.6
          Copyright (c) 1993. - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:      January 14, 2006, 14:44:31 ; Search time 398 Seconds
                                  (without alignments)
                                  3056.409 Million cell updates/sec

Title:           US-10-618-976-1
Perfect score:   1503
Sequence:        1 atgcaactccaacgcgcttt.........ccgaggccagccagcgcgct 1503

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        6038814 seqs, 404674181 residues

Total number of hits satisfying chosen parameters:  12077628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published Applications NA New:*
                 1: /cgn2_6/ptcdata/2/pubpna/US08_NEW_PUB.seq:*
                 2: /cgn2_6/ptcdata/2/pubpna/US06_NEW_PUB.seq:*
                 3: /cgn2_6/ptcdata/2/pubpna/US07_NEW_PUB.seq:*
                 4: /cgn2_6/ptcdata/2/pubpna/PCT_NEW_PUB.seq:*
                 5: /cgn2_6/ptcdata/2/pubpna/US09_NEW_PUB.seq:*
                 6: /cgn2_6/ptcdata/2/pubpna/US10_NEW_PUB.seq:*
                 7: /cgn2_6/ptcdata/2/pubpna/US11_NEW_PUB.seq:*
                 8: /cgn2_6/ptcdata/2/pubpna/US11_NEW_PUB.seq2:*
                 9: /cgn2_6/ptcdata/2/pubpna/US11_NEW_PUB.seq3:*
                10: /cgn2_6/ptcdata/2/pubpna/US60_NEW_PUB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	Length	ID	Description
1	43	2.9	88421	7	US-11-205-109-1	Sequence 1, Appli
2	40.2	2.7	1647	7	US-11-137-465-14	Sequence 14, Appl
3	39.6	2.6	761	6	US-10-497-135-18	Sequence 18, Appl
4	39	2.6	88421	7	US-11-205-109-1	Sequence 1, Appli
5	38.6	2.6	3618	7	US-11-136-527-294	Sequence 294, App
6	38.2	2.5	1400	7	US-11-136-527-4370	Sequence 4370, Ap
7	38.2	2.5	6021	7	US-11-136-527-274	Sequence 274, App
8	38	2.5	2749	7	US-11-000-688-100	Sequence 100, App
9	37.6	2.5	2149	7	US-11-078-189-6	Sequence 6, Appli
10	37.2	2.5	10968	7	US-11-075-185-35	Sequence 35, Appl
11	37.2	2.5	78869	7	US-11-075-185-1	Sequence 1, Appli
12	37.2	2.5	116856	7	US-11-143-980-1	Sequence 1, Appli
13	37.2	2.5	168516	7	US-11-121-086-3	Sequence 3, Appli
14	36.8	2.4	10524	7	US-11-075-185-37	Sequence 37, Appl
15	36.6	2.4	1987	7	US-11-128-061-384	Sequence 384, App
16	36.4	2.4	767	6	US-10-497-135-17	Sequence 17, Appl
17	36.4	2.4	7006	6	US-10-821-234-218	Sequence 218, App
18	35.6	2.4	864	7	US-11-179-411-26	Sequence 26, Appl
19	35.6	2.4	864	7	US-11-175-766-26	Sequence 26, Appl
20	35.6	2.4	1974	6	US-10-821-234-34	Sequence 34, Appl
21	35.6	2.4	11070	7	US-11-075-185-34	Sequence 34, Appl
22	35.6	2.4	37507	6	US-10-522-037-2	Sequence 2, Appli
23	35.4	2.4	2023	6	US-10-995-561-55	Sequence 55, Appl
c 24	35.4	2.4	2391	7	US-11-136-527-1926	Sequence 1926, Ap
25	35.4	2.4	11684	6	US-10-995-561-498	Sequence 498, App
26	35.4	2.4	12591	6	US-10-995-561-13415	Sequence 13415, A
27	35.4	2.4	14023	6	US-10-995-561-13221	Sequence 13221, A
c 28	35.4	2.4	199130	6	US-10-995-561-13233	Sequence 13233, A
29	35.2	2.3	930	6	US-10-524-647-127	Sequence 127, App
30	35.2	2.3	2570	7	US-11-136-527-3380	Sequence 3380, Ap
31	35	2.3	1290	7	US-11-136-527-2457	Sequence 2457, Ap
32	35	2.3	1290	7	US-11-136-527-6553	Sequence 6553, Ap
c 33	35	2.3	116856	7	US-11-143-980-1	Sequence 1, Appli
34	34.8	2.3	2025	7	US-11-167-048-2	Sequence 2, Appli
35	34.8	2.3	2736	6	US-10-858-730-38	Sequence 38, Appl
36	34.8	2.3	5679	7	US-11-075-185-36	Sequence 36, Appl
c 37	34.8	2.3	155515	7	US-11-112-908-42	Sequence 42, Appl
c 38	34.8	2.3	159660	7	US-11-112-908-43	Sequence 43, Appl
c 39	34.8	2.3	177623	7	US-11-112-908-41	Sequence 41, Appl
c 40	34.6	2.3	609	6	US-10-467-657-8011	Sequence 8011, Ap
41	34.6	2.3	690	6	US-10-858-730-156	Sequence 156, App
42	34.6	2.3	6941	6	US-10-432-483-49	Sequence 49, Appl
43	34.6	2.3	179597	7	US-11-121-086-91	Sequence 91, Appl
44	34.4	2.3	489	7	US-11-120-308-5	Sequence 5, Appli
45	34.4	2.3	543	7	US-11-120-308-11	Sequence 11, Appl

ALIGNMENTS

```
RESULT 1
US-11-205-109-1
; Sequence 1, Application US/11205109
; Publication No. US2005028764lA1
; GENERAL INFORMATION:
;  APPLICANT: Farnet, Chris
;  APPLICANT: Zazopoulos, Emmanuel
;  APPLICANT: Staffa, Alfredo
;  TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS
;  FILE REFERENCE: 3002-2US
;  CURRENT APPLICATION NUMBER: US/11/205,109
;  CURRENT FILING DATE: 2005-08-17
;  PRIOR APPLICATION NUMBER: US 60/239,924
;  PRIOR FILING DATE: 2001-10-15
;  PRIOR APPLICATION NUMBER: US 60/239,924
;  PRIOR FILING DATE: 2000-10-13
;  NUMBER OF SEQ ID NOS: 46
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO. 1
;  LENGTH: 88421
;  TYPE: DNA
;  ORGANISM: Actinoplanes sp.
;  FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (2077)..(3078)
;    OTHER INFORMATION: ORF 1; positive strandedness
;  FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (3118)..(4032)
;    OTHER INFORMATION: ORF 2; positive strandedness
;  FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (4038)..(5048)
;    OTHER INFORMATION: ORF 3; positive strandedness
;  FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (6665)..(5814)
;    OTHER INFORMATION: ORF 4; negative strandedness
;  FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (7703)..(6693)
;    OTHER INFORMATION: ORF 5; negative strandedness
;  FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (9464)..(8130)
;    OTHER INFORMATION: ORF 6; negative strandedness
```

FEATURE:
NAME/KEY: misc_feature
LOCATION: (9691)..(10761)
OTHER INFORMATION: ORF 7; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (12751)..(10829)
OTHER INFORMATION: ORF 8; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (13617)..(12802)
OTHER INFORMATION: ORF 9; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (15203)..(13614)
OTHER INFORMATION: ORF 10; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (15591)..(15863)
OTHER INFORMATION: ORF 11; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (15880)..(19035)
OTHER INFORMATION: ORF 12; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (19032)..(39713)
OTHER INFORMATION: ORF 13; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (39713)..(65800)
OTHER INFORMATION: ORF 14; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65826)..(66530)
OTHER INFORMATION: ORF 15; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (66546)..(67370)
OTHER INFORMATION: ORF 16; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (67384)..(70059)
OTHER INFORMATION: ORF 17; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (70099)..(70662)
OTHER INFORMATION: ORF 18; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (70659)..(71906)
OTHER INFORMATION: ORF 19; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (73439)..(71964)
OTHER INFORMATION: ORF 20; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (74216)..(73563)
OTHER INFORMATION: ORF 21; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (75424)..(74213)
OTHER INFORMATION: ORF 22; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (75535)..(76464)
OTHER INFORMATION: ORF 23; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (78110)..(76449)
OTHER INFORMATION: ORF 24; negative strandedness
FEATURE:

NAME/KEY: misc_feature
LOCATION: (79864)..(78107)
OTHER INFORMATION: ORF 25; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (81624)..(79861)
OTHER INFORMATION: ORF 26; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (81909)..(81682)
OTHER INFORMATION: ORF 27; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (82346)..(82062)
OTHER INFORMATION: ORF 28; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (82587)..(84446)
OTHER INFORMATION: ORF 29; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (84481)..(85548)
OTHER INFORMATION: ORF 30; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (85556)..(86845)
OTHER INFORMATION: ORF 31; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (87372)..(86803)
OTHER INFORMATION: ORF 32; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (87494)..(88420)
OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-11-205-109-1

```
Query Match            2.9%;  Score 43;  DB 7;  Length 88421;
Best Local Similarity   50.7%;  Pred. No. 0.92;  Indels 0;  Gaps  0;
Matches  103;  Conservative   0;  Mismatches 100;  Indels  0;

Qy      1267 GGCATCCGTTGTTGTCGGAATACGAGGTCGCCGCCTTTGATCGAGCTGGCCTTCAAA 1326
                  ||  ||||  | ||||||||    |      |||||||||||||
Db     16309 GGCAGCCTTCGGCATTGTCGCCTCCCGGTGGCCGCGTCTATTCCGCCGCTCGCGGCCCGT 16368

Qy      1327 AGCCGGCCAAGCTGGCTATTGTGCCTTTGCAGGACGTGCTGGGCGCCGGCCCCGAGGCC 1386
                  |||     ||  |    ||  | |   ||||||   ||||||| ||||||||| |||
Db     16369 GGTGACGTCCGGCCGGGTGGCGCTGCGGCCGCTGTCGGTGTTGTCGTCGCGGCCCGACGCC 16428

Qy      1387 CGCATGAACTTCCCCGACGGCTGGGGACAACTGGCGGTGGCGCTACGGCCGAAGGCGAC 1446
                  ||  | |   ||||   |||||| |  | |||||  |||| |||  ||||||| |||
Db     16429 TACGAGCGGCTCCGGCGACCGGGACCCGGAGCACTGGCCGCTCCCGCGCTGCGCGGGC 16488

Qy      1447 CTCGAGCCCGGTCTGGCCGCGGG 1469
                  || ||  ||| ||| ||||||
Db     16489 CTGCCCGCCGAGCTGCTCGCGGG 16511
```

```
      ;   PRIOR APPLICATION NUMBER: US/10/239,663
      ;   PRIOR FILING DATE: 2002-09-24
      ;   PRIOR APPLICATION NUMBER: PCT/US01/09226
      ;   PRIOR FILING DATE: 2001-03-22
      ;   PRIOR APPLICATION NUMBER: 60/192,158
      ;   PRIOR FILING DATE: 2000-03-24
      ;   PRIOR APPLICATION NUMBER: 60/192,668
      ;   PRIOR FILING DATE: 2000-03-27
      ;   PRIOR APPLICATION NUMBER: 60/200,166
      ;   PRIOR FILING DATE: 2000-04-27
      ;   NUMBER OF SEQ ID NOS: 66
      ;   SOFTWARE: FastSEQ for Windows Version 3.0
      ;   SEQ ID NO 14
      ;     LENGTH: 1647
      ;     TYPE: DNA
      ;     ORGANISM: Homo sapiens
      US-11-137-465-14

      Query Match            2.7%;    Score 40.2; DB 7;  Length 1647;
      Best Local Similarity  48.5%;   Pred. No. 2.5;
      Matches  111;  Conservative  0;  Mismatches  118;  Indels  0;  Gaps  0;

      Qy    1254 CCTGGCCCGCTATGGCATCCGTTGTTTGTCGAATACGAGGTCGCGGCGCTTTGATCGA 1313
                 ||||| ||| || | |   |||| |  | || | ||  || || || ||  | ||||
      Db     336 CGTGTGCGCTGCCGCGACCGCTGTGAGAAGGAGCCCAGCTTCACCTGCCGTCGGACGG 395

      Qy    1314 GCTGGCCTTCAAAAGCCCGGCCCAAGCTGGCTATTGTGCCTTTGCAGGACGTGCTGGGGCT 1373
                 ||| |||    | || |||| | ||||  ||  ||  ||| | || |  | ||  ||||||
      Db     396 CCTCACCTACTACAACCGCTGCTATATGGACCGCCGAGCCTCGCCTGCGGGGCCTGCACCT 455

      Qy    1374 GGGCCCCGAGGCCCGCATGAACTTCCCCGGACGGCTGGGGACGAACTGGCGTGGCCGCTA 1433
                 | | | |      |  || |||| || | |  ||| |  | |    ||| | | |  ||
      Db     456 CCACATCGTGCCCTGCCTGCAAGCACGTGCTCAGCTGGCCGCCGCCCAGCAGCCCGGCCGGA 515

      Qy    1434 CGCGGAGGCGACCTCGAGCCCGGCTGTCGGCCGCCGGGGACTGCGGGCCCTG 1482
                 |   |||   | ||    | ||| |   || |||    |   | | |||||
      Db     516 GACCACTGCCCCGCCCCACACCTGGGGCGGCCGCGCCGTGCTCCTGCCCCTG 564


      RESULT 3
      US-10-497-135-18
      ;   Sequence 18, Application US/10497135
      ;   Publication No. US20050272213A1
      ;   GENERAL INFORMATION:
      ;     APPLICANT: Biotica Technology Ltd
      ;     APPLICANT: Gregory, Matthew A
      ;     APPLICANT: Gaisser, Sabine
      ;     APPLICANT: Petkovic, Hrvoje
      ;     APPLICANT: Moss, Steven
      ;   TITLE OF INVENTION: Production of Polyketides and Other Natural Products
      ;   FILE REFERENCE: 4408-P03444US00
      ;   CURRENT APPLICATION NUMBER: US/10/497,135
      ;   CURRENT FILING DATE: 2004-05-26
      ;   PRIOR APPLICATION NUMBER: PCT/GB03/003230
      ;   PRIOR FILING DATE: 2003-07-16
      ;   PRIOR APPLICATION NUMBER: GB0216509.0
      ;   PRIOR FILING DATE: 2002-07-16
      ;   PRIOR APPLICATION NUMBER: GB0224922.5
      ;   PRIOR FILING DATE: 2002-10-25
      ;   NUMBER OF SEQ ID NOS: 63
      ;   SOFTWARE: PatentIn version 3.2
      ;   SEQ ID NO 18
      ;     LENGTH: 761
      ;     TYPE: DNA
      ;     ORGANISM: Streptomyces hygroscopicus
      US-10-497-135-18

      Query Match            2.6%;    Score 39.6; DB 6;  Length 761;
      Best Local Similarity  47.7%;   Pred. No. 3.1;
      Matches  148;  Conservative  0;  Mismatches  159;  Indels  3;  Gaps  1;

      Qy    1117 GAGGACAACGCGCCTTTTTGCCCCACAACTACCCCGCCACGCGGCAATGGTGGTGGTACAGC 1176
                 |||||| ||  ||   |     ||   ||  ||  | | |||  | | | ||| |||  ||
```

```
      Db     441 GAGGACGTCGAGCTGCACGGCCAGACCATCAGCGGCGACAACGTGATCCGCTCTAC 500
                 |||||   |||   |  |  |  |   |||  ||||| || ||  |  ||  |||
      Qy    1177 GGAACCCAGCAGACGACGCACCACCTGGATGGTTCCGCCGAGGCCGAGCGG 1236
                 ||   ||  |  |   |  |  || || ||  | || | |||  | || |||
      Db     501 TCGACGGCCAACCGCCGACCCCGGCGTCTTCGCCGACCCCGACACGTTCGACATCACGCGT 560
                 |||| |   || | | |  ||| || |  | ||  | |||| || | | ||| |
      Qy    1237 GCCTTCATGCGGCGGCTACCTGGCCGTATGGCATCCGTTGTTGTCGGAATACGAGGTC 1296
                 ||   || | | | | | || || ||  || ||||||| || | | ||  |  ||
      Db     561 AAGCCCGAACAACACTTCGCTTTCGGGGTACGGACGCATCACGCCGTGCCCGGGCAGCACCTC 620
                 |||  || |  | | |  ||  |||||  | |  | | | |   || || | ||| |
      Qy    1297 GCGGGCGCTTTGATCGAGCTGGCCTTCAAAAGCCCGGCCCAAGC---TGGCTATTGTGCCT 1353
                 | ||   |  | |  | | ||  ||| |   | ||  || ||       ||  ||  ||
      Db     621 GCCCGCGTGTTGATCAAGGTCGCCACCGTGCCGCCGCCGTTCGGAGCGCTTCCCCGGACCGGA 680
                 |||  | | | ||| |  | || || |  | ||| || |   |||  |  |||  ||
      Qy    1354 TTGCAGGACGTGCTGGGGCTGGGCTGGGGCGTGGGGCCCGCATGAACTTCCCCGGG 1413
                 || |||||| ||| ||||  | ||  || |   || |||  |||  ||  ||||||
      Db     681 CTGGCGGCGCACGTGCCCATGAACGAGAGGGTCTCGGCCTCGTTCAGCCCGGCCGCGAGCTCCGG 740
                 |||  |||| | ||  || ||| || |  | ||| ||| |  ||| |  |  ||| |||||
      Qy    1414 GACAACTCGG 1423
                 | || ||||
      Db     741 GTCCACCTGG 750


      RESULT 4
      US-11-205-109-1/c
      ;   Sequence 1, Application US/11205109
      ;   Publication No. US20050287641A1
      ;   GENERAL INFORMATION:
      ;     APPLICANT: Farnet, Chris
      ;     APPLICANT: Zazopoulos, Emmanuel
      ;     APPLICANT: Staffa, Alfredo
      ;   TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS
      ;   FILE REFERENCE: 3002-2US
      ;   CURRENT APPLICATION NUMBER: US/11/205,109
      ;   CURRENT FILING DATE: 2005-08-17
      ;   PRIOR APPLICATION NUMBER: US/09/976,059
      ;   PRIOR FILING DATE: 2001-10-15
      ;   PRIOR APPLICATION NUMBER: US 60/239,924
      ;   PRIOR FILING DATE: 2000-10-13
      ;   NUMBER OF SEQ ID NOS: 46
      ;   SOFTWARE: PatentIn version 3.0
      ;   SEQ ID NO 1
      ;     LENGTH: 88421
      ;     TYPE: DNA
      ;     ORGANISM: Actinoplanes sp.
      ;   FEATURE:
      ;     NAME/KEY: misc_feature
      ;     LOCATION: (2077)..(3078)
      ;     OTHER INFORMATION: ORF 1; positive strandedness
      ;   FEATURE:
      ;     NAME/KEY: misc_feature
      ;     LOCATION: (3118)..(4032)
      ;     OTHER INFORMATION: ORF 2; positive strandedness
      ;   FEATURE:
      ;     NAME/KEY: misc_feature
      ;     LOCATION: (4038)..(5048)
      ;     OTHER INFORMATION: ORF 3; positive strandedness
      ;   FEATURE:
      ;     NAME/KEY: misc_feature
      ;     LOCATION: (6665)..(5814)
      ;     OTHER INFORMATION: ORF 4; negative strandedness
      ;   FEATURE:
      ;     NAME/KEY: misc_feature
      ;     LOCATION: (7703)..(6693)
      ;     OTHER INFORMATION: ORF 5; negative strandedness
      ;   FEATURE:
      ;     NAME/KEY: misc_feature
      ;     LOCATION: (9464)..(8130)
      ;     OTHER INFORMATION: ORF 6; negative strandedness
      ;   FEATURE:
      ;     NAME/KEY: misc_feature
      ;     LOCATION: (9691)..(10761)
      ;     OTHER INFORMATION: ORF 7; positive strandedness
```

FEATURE:
NAME/KEY: misc_feature
LOCATION: (12751)..(10829)
OTHER INFORMATION: ORF 8; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (13617)..(12802)
OTHER INFORMATION: ORF 9; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (15203)..(13614)
OTHER INFORMATION: ORF 10; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (15591)..(15863)
OTHER INFORMATION: ORF 11; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (15880)..(19035)
OTHER INFORMATION: ORF 12; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (19032)..(39713)
OTHER INFORMATION: ORF 13; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (39713)..(65800)
OTHER INFORMATION: ORF 14; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65826)..(66530)
OTHER INFORMATION: ORF 15; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (66546)..(67370)
OTHER INFORMATION: ORF 16; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (67384)..(70059)
OTHER INFORMATION: ORF 17; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (70099)..(70662)
OTHER INFORMATION: ORF 18; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (70659)..(71906)
OTHER INFORMATION: ORF 19; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (73439)..(71964)
OTHER INFORMATION: ORF 20; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (74216)..(73563)
OTHER INFORMATION: ORF 21; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (75424)..(74213)
OTHER INFORMATION: ORF 22; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (75535)..(76464)
OTHER INFORMATION: ORF 23; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (78110)..(76449)
OTHER INFORMATION: ORF 24; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (79864)..(78107)
OTHER INFORMATION: ORF 25; negative strandedness
FEATURE:

FEATURE:
NAME/KEY: misc_feature
LOCATION: (81624)..(79861)
OTHER INFORMATION: ORF 26; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (81909)..(81682)
OTHER INFORMATION: ORF 27; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (82346)..(82062)
OTHER INFORMATION: ORF 28; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (82587)..(84446)
OTHER INFORMATION: ORF 29; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (84481)..(85548)
OTHER INFORMATION: ORF 30; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (85556)..(86845)
OTHER INFORMATION: ORF 31; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (87372)..(86803)
OTHER INFORMATION: ORF 32; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (87494)..(88420)
OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only

US-11-205-109-1

Query Match   2.6%;   Score 39;   DB 7;   Length 88421;
Best Local Similarity   48.1%;   Pred. No. 7.8;
Matches   111;   Conservative   0;   Mismatches   120;   Indels   0;   Gaps   0;

Qy   1133 TGCCCCACAACTACCCGCGCACGGCAATGTGGTGGTGTACAGCGGAACCCACGACAACG 1192
                |||||| || ||||| |||| ||| || ||
Db 54507 TGCACGATCGTCACGGGCGTCGGGTCGGAGGCGGGCGACCCGGCCCGGTCGGCACC 54448

Qy   1193 ACACCACCCTGGGATGGTTCCGCACCGCCGGCGGCGCGAGGCCCGAGCCGGAGCG 1252
                ||||| ||| ||||| |||||| |||||
Db 54447 ACGGCGTCCTCGGCGACGACACCTCCGGCGCCGCCGACCAGCGCGCACGTAGAGCTTCCAGG 54388

Qy   1253 ACCTGGCCCGCTATGGCATCCGTTGTTTGTCGGAATACGAGGTCGCGGGCGCTTTGATCG 1312
                ||||| |||||| || ||| |||||| |||||
Db 54387 ACGTGGATGTCGAAGGCGACCCGGTGACCGGGCAGCAGCGTCGGCGCGCCCAGCGCG 54328

Qy   1313 AGCTGGCCTTCAAAAGCCGGCCAAGCTGGCTATTGTGCCTTTGCAGGACG 1363
                ||||||| || || ||| ||||
Db 54327 AGCCGGTCGTCATCGCCGCCACGAAATTCGTGAGGGCCCGTGCGGGCACG 54277

RESULT 5
US-11-136-527-294
; Sequence 294, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 294
; LENGTH: 3618
; TYPE: DNA
; ORGANISM: Rattus norvegicus

```
;  FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (3524)..(3525)
;    OTHER INFORMATION: n is a, c, g, or t
US-11-136-527-294

Query Match       2.6%;     Score 38.6;  DB 7;  Length 3618;
Best Local Similarity   50.3%;  Pred. No. 6.4;
Matches   95;  Conservative   0;  Mismatches   94;  Indels   0;  Gaps   0;

Qy   366  GGCCCAGGATAAGACCCGACTGGAAGCCTTATCGAGGCCGAGCGCTTCTGGCTGGAAGA  425
               |||||  |||  |||
Db   366  GGAAAAGATTGAGAAGGAGCTGGAGGCCGTGTGCCAGGACGTGCTGCTGAGCCTGCTGGACAA  425

Qy   426  CTTATGGCTCTTTATGGCCCTCAAGACACCCGGTTTGACGGCAAGCCTGGAACGAGTCGAG  485
               |||||  |||||
Db   426  CTACCTGATCAAGAACTGCAGCCGAGACCCAGTACGAGCGAGCAAGGTGTTCTACCTGAAGAT  485

Qy   486  CCCCGAGCTGCGCGACCGTGAACCGG  545
               |||||||
Db   486  GAAAGGGGACTATTACCGCTGGCTACCTGGCTGAAGTGGCCACTGGGGGAGAAAAGGGCGACCGT  545

Qy   546  GGAGGTGGC  554
               ||  |||
Db   546  GGTGGAGTC  554


RESULT 6
US-11-136-527-4370
; Sequence 4370, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
;    APPLICANT: Wyeth
;    APPLICANT: Mounts, William M
;    TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
;    FILE REFERENCE: 031896-041000 (AM101086)
;    CURRENT APPLICATION NUMBER: US/11/136,527
;    CURRENT FILING DATE: 2005-05-25
;    PRIOR APPLICATION NUMBER: US 60/574,294
;    PRIOR FILING DATE: 2005-05-26
;    NUMBER OF SEQ ID NOS: 362830
;    SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4370
;    LENGTH: 1400
;    TYPE: DNA
;    ORGANISM: Rattus norvegicus
US-11-136-527-4370

Query Match       2.5%;     Score 38.2;  DB 7;  Length 1400;
Best Local Similarity   39.4%;  Pred. No. 7;
Matches   98;  Conservative   27;  Mismatches   124;  Indels   0;  Gaps   0;

Qy   400  GAGGCCCAGCGCGTTCTGGCTGGAAGACTATGCCCTTTATGCCCTCAAGACACCCGGTTT  459
               |||  ||||||  :||
Db    11  GAGAACAAGAACCTCCAGGAGGAGATCTCAGACTGACTGCGACCTGACCGTGCCGGCTGCCCTG  519

Qy   460  GACGGCAAGCCCCTGGAACGAGTGGAGCGCGAGAGGTCCCGAGGAGGTGGCCCTGCCCTG  519
               :::  :||||  ||||  ||||  :|  :||
Db    71  WARARBRTSCACGAGCTGGAGAAGATCCGMAARCARCTGGAGGYKGADMWKCTSSWMSTG  130

Qy   520  GCCAGGGCCCGTGAGGAGCTGGCCGAGGAGGTGGCCCTTTACGAGTGGATTCAGTGGCTT  579
               :||  |||||  |||||
Db   131  CMGTCAGCCCTGGAGGAGGYTGAGGCGTGCGAGGGCCAAGATCCTCCGA  190

Qy   580  TTTTATCTGGAATGGGGCCAGACCAAGGCCTATGCCGAATCCAAGGGGATTCAGATTATC  639
               :|  ||||||  :|||
Db   191  GCYCAGCTGGAGTTCAMCCAGATYAAGGSAGAGATYGAAAGGAAGCTGCGGAGGAADGAC  250

Qy   640  GGCGATATG  648
               ||  |||
Db   251  GAGGAGATK  259
```

```
RESULT 7
US-11-136-527-274
; Sequence 274, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
;    APPLICANT: Wyeth
;    APPLICANT: Mounts, William M
;    TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
;    FILE REFERENCE: 031896-041000 (AM101086)
;    CURRENT APPLICATION NUMBER: US/11/136,527
;    CURRENT FILING DATE: 2005-05-25
;    PRIOR APPLICATION NUMBER: US 60/574,294
;    PRIOR FILING DATE: 2005-05-26
;    NUMBER OF SEQ ID NOS: 362830
;    SOFTWARE: PatentIn version 3.2
; SEQ ID NO 274
;    LENGTH: 6021
;    TYPE: DNA
;    ORGANISM: Rattus norvegicus
US-11-136-527-274

Query Match       2.5%;     Score 38.2;  DB 7;  Length 6021;
Best Local Similarity   39.4%;  Pred. No. 8.4;
Matches   98;  Conservative   27;  Mismatches   124;  Indels   0;  Gaps   0;

Qy   400  GAGGCCCAGCGCGTTCTGGCTGGAAGACTATGCCCTTTATGCCCTCAAGACACCCGGTTT  459
               |||  ||||||  :||
Db  4632  GAGAACAAGAACCTCCAGGAGGAGATCTCAGACTGACTGCGACCTGACCGTGCCGGCTGCCCTG  4691

Qy   460  GACGGCAAGCCCCTGGAACGAGTGGAGCGCGAGAGGTCCCGAGGAGGTGGCCCTGCCCTG  519
               :::  :||||  ||||  ||||  :|  :||
Db  4692  WARARBRTSCACGAGCTGGAGAAGATCCGMAARCARCTGGAGGYKGADMWKCTSSWMSTG  4751

Qy   520  GCCAGGGCCCGTGAGGAGCTGGCCGAGGAGGTGGCCCTTTACGAGTGGATTCAGTGGCTT  579
               :||  |||||  |||||
Db  4752  CMGTCAGCCCTGGAGGAGGYTGAGGCGTGCGAGGGCCAAGATCCTCCGA  4811

Qy   580  TTTTATCTGGAATGGGGCCAGACCAAGGCCTATGCCGAATCCAAGGGGATTCAGATTATC  639
               :|  ||||||  :|||
Db  4812  GCYCAGCTGGAGTTCAMCCAGATYAAGGSAGAGATYGAAAGGAAGCTGCGGAGGAADGAC  4871

Qy   640  GGCGATATG  648
               ||  |||
Db  4872  GAGGAGATK  4880


RESULT 8
US-11-000-688-100/c
; Sequence 100, Application US/11000688
; Publication No. US20050287544A1
; GENERAL INFORMATION:
;    APPLICANT: BERTUCCI, Francois
;    APPLICANT: HOULGATTE, Remi
;    APPLICANT: BIRNBAUM, Daniel
;    TITLE OF INVENTION: GENE EXPRESSION PROFILING OF COLON CANCER WITH DNA ARRAYS
;    FILE REFERENCE: 1423-R-03
;    CURRENT APPLICATION NUMBER: US/11/000,688
;    CURRENT FILING DATE: 2004-12-01
;    PRIOR APPLICATION NUMBER: US 60/525,987
;    PRIOR FILING DATE: 2003-12-01
;    NUMBER OF SEQ ID NOS: 1596
;    SOFTWARE: PatentIn version 3.2
; SEQ ID NO 100
;    LENGTH: 2749
;    TYPE: DNA
;    ORGANISM: Artificial Sequence
;    FEATURE:
;    OTHER INFORMATION: Description of Artificial sequences:primer
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (1)..(2749)
;    OTHER INFORMATION: colony stimulating factor 1
;    OTHER INFORMATION: (macrophage) (CSF1) gene.
```

US-11-000-688-100

Query Match   2.5%;   Score 38;   DB 7;   Length 2749;
Best Local Similarity   51.9%;   Pred. No. 8.5;
Matches 109;   Conservative   0;   Mismatches 100;   Indels 1;   Gaps 1;

Qy 1286 AATACGAGAGTCGCGGCGCGCTTGATCGAGCTGGCCTTCAAAAGCCGGCCAAGCTGGCTA 1345
Db   304 AGTACTCCGACACCTCCTCGGTGATACTCCTGCTCGCCAGGAGACAGACCAACAACAGCA 245

Qy 1346 TTGTGCCTTTGCAGGACGTGCTGGGGCTGGGCCCCGGAGGC-CCGCATGAACTTCCCCGGA 1404
Db   244 GGGAGCCCGACAACTGTCGTGGGAGGGCAGCGCCCGCGGGGTCATACGGG 185

Qy 1405 CGGCTGGGGGACAACTGGGCGTGGCGCTACGCCGAAGGGCGACCTCGAGCCCGGTCTGGCC 1464
Db   184 CAGCTGGGGTCCCGCGCGGGGCGCGTCGGGCCCCTCGCTCGCGCTCGCTGGCTGC 125

Qy 1465 GCGGGACTGCGGGCCCTGGCCGCCGAGGCCAGC 1494
Db   124 TGCGGAGTGGGCGCGCCCCCGAGGCCGGC 95

RESULT 9
US-11-078-189-6
; Sequence 6, Application US/11078189
; Publication No. US2005027716A1
; GENERAL INFORMATION:
; APPLICANT: Bachinger, Hans Peter
; APPLICANT: Vranka, Janice
; TITLE OF INVENTION: PROLYL 3-HYDROXYLASES
; FILE REFERENCE: 08062-020001
; CURRENT APPLICATION NUMBER: US/11/078,189
; CURRENT FILING DATE: 2005-03-11
; PRIOR APPLICATION NUMBER: US 60/552,409
; PRIOR FILING DATE: 2004-03-11
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 2149
; TYPE: DNA
; ORGANISM: Mus musculus
US-11-078-189-6

Query Match   2.5%;   Score 37.6;   DB 7;   Length 2149;
Best Local Similarity   47.8%;   Pred. No. 10;
Matches 109;   Conservative   0;   Mismatches 119;   Indels 0;   Gaps 0;

Qy 1225 GAGGCCGAGGGGGCCTTCATGCGGGCCTACCTGGCTGGGGCTGGGGCCTGGGCTATGGGAT...GCCGGCCATGGGTCG 1284
Db   250 GCGGGCCGAGCCGGCGGGCCGCGGCGCTTGGGGCTCCCTCCCAGCTCCTCCGAGCCCCACATCCTGTCTCC 309

Qy 1285 GAATACAGAGGTCGCGGCGCTTTGATCGAGCTGGCCTTCAAAAGCCGGCCAAGCTGGCT 1344
Db   310 GGGCCCCGGGGTCTGGGAACCGCTGCTCCTGCGCGCGCTGGGGCGGGGCGCCGGAGTGGCCTG 369

Qy 1345 ATTGTGCCTTTGCAGGACGTGCTGGGGCTGGGGCCCCGGAGGCCCGCATGAACTTCCCCGGA 1404
Db   370 ACCCAGTGCGCGGTGCGGAGGCTGGGCGGCGCGGGGGGCGCGGGGCGCCTTCCCGCTGGGGGAGC 429

Qy 1405 CGGCTGGGGGGACAACTGGGCGCGCCTGGCCGCCTACGCCGAAGGGCGACCTCGAG 1452
Db   430 GCGCTGCGGCGGATGCGCTTCCCCGCGCGGGGGAGCGCCTTACAACTACCTGCAG 477

RESULT 10
US-11-075-185-35
; Sequence 35, Application US/11075185
; Publication No. US2005026643A1
; GENERAL INFORMATION:
; APPLICANT: REEVES, CHRISTOPHER D
; APPLICANT: JULIEN, BRYAN
; APPLICANT: REID, RALPH

TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
; FILE REFERENCE: 010099.03
; CURRENT APPLICATION NUMBER: US/11/075,185
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/551,103
; PRIOR FILING DATE: 2004-03-08
; PRIOR APPLICATION NUMBER: US 60/568,290
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 35
; LENGTH: 10968
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-11-075-185-35

Query Match   2.5%;   Score 37.2;   DB 7;   Length 10968;
Best Local Similarity   45.7%;   Pred. No. 16;
Matches 129;   Conservative   0;   Mismatches 153;   Indels 0;   Gaps 0;

Qy 1183 CACGACAACGACACCACCTGGGATGGTTCCGCACCGGCCGGAGGCCGAGGCGGCCTTC 1242
Db   7090 CGGCGCCCTGTACGCCACCCCGTCTTCCGCGGCCGGGCGTCTGCGCGCTGCGCCGAG 7149

Qy 1243 ATGCGGGCCTACCTGGCCCGCTATGCGCATCCGTTGTTGTCGGAATACGAGGTCGCGGGC 1302
Db   7150 CTCGACCGGCCACCTCGACCGGCCCCTCATGAGCGTCCATGAGGGTCCTCTTCGCCGACGCGCCGGCTCCGAG 7209

Qy 1303 GCTTTGATCGAGCTGGCCTTCAAAAGCCCGGCCAAGCTGGCTATTGTGCCTTTGCGCTCTCGAGGAC 1362
Db   7210 GCCGGCGGCGCTGCTCGACCAGACAGCGTGGGCTCAGCCGCCCCTTCCCCGACGGCTGGGGGACAACTGG 7269

Qy 1363 GTGCTGGGGCTGGGCCCCGGAGGCCGCATGATGAACTTCCCCGACGGCTGGGGGACAACTGG 1422
Db   7270 GCCCCTACCGGCCAGTGGGATGCCTGGGGCCCGGGGCCTGCGCCCCACAGCTGCTCGGCCACAGC 7329

Qy 1423 GCGTGCGCTACGCCGAAGGCGACCTCGAGCCCGGTCTGGCC 1464
Db   7330 ATCGGCGAGCTCGCGCGCCCGCCCACATCGCCGCGGTGCTCGAC 7371

RESULT 11
US-11-075-185-1
; Sequence 1, Application US/11075185
; Publication No. US2005026643A1
; GENERAL INFORMATION:
; APPLICANT: REEVES, CHRISTOPHER D
; APPLICANT: JULIEN, BRYAN
; APPLICANT: REID, RALPH
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
; FILE REFERENCE: 010099.03
; CURRENT APPLICATION NUMBER: US/11/075,185
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/551,103
; PRIOR FILING DATE: 2004-03-08
; PRIOR APPLICATION NUMBER: US 60/568,290
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 78869
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-11-075-185-1

Query Match   2.5%;   Score 37.2;   DB 7;   Length 78869;
Best Local Similarity   45.7%;   Pred. No. 20;
Matches 129;   Conservative   0;   Mismatches 153;   Indels 0;   Gaps 0;

Qy   1183 CACGACAACGACACCACCTGGGATGGTTCCGCACCGGCCGGAGGCCGAGGCGGCCTTC 1242
Db 22357 CGGCGCCCTGTACGCCACCCCGTCTTCCGCGGCCGGCTGCGACGCGCCTGCGCCGAG 22416

```
Qy   1243  ATGCGGGCTACCTGGCCGCCGCTATGGCATCCGTTGTTGTCGGAATACGAGGTCGCGGGC 1302
                 |||||        ||||                |||||     ||||
Db  22417  CTCGACCGGCCACCTCGACCGGCCCCCATGAGCGTCCTTCTTCGCCGACGGCCGGCTCCGAG 22476

Qy   1303  GCTTTGATCGAGCTGGCCTTCAAAAGCCCGGCCAAGCTGGCTATTGTGCCTTTGCAGGAC 1362
                |||||        |||||                |||||
Db  22477  GCCGCGGCGGCTGCTGCTGCGACCAGACAGCCCTGGGCTCAGCCGCCGCCTCTTCGCTCGAGGTG 22536

Qy   1363  GTGCTGGGGCTGGGCCCCGAGGCCCGCATGAACTTCCCCGACGGCGCTGGGGGACAACTGG 1422
                |||||        ||||                |||||     ||||
Db  22537  GCCCTCTTACCGCCAGTGGGATGCCTGGGGCCTGCGCCCGAGCTGCTGCTCCGGCCACCAGC 22596

Qy   1423  GCGTGGCGCTACCGCGAAGGCGACCTCGAGCCCGGTCTGGCC 1464
                |||||        |||||
Db  22597  ATCGGGCGCTCGCCGCGCCGCCACATCGCCGCGTGCTCGAC 22638


RESULT 12
US-11-143-980-1
; Sequence 1, Application US/11143980
; Publication No. US20050272133A1
; GENERAL INFORMATION:
; APPLICANT: He, Min
; APPLICANT: Hucul, John
; APPLICANT: Haltli, Bradley A.
; APPLICANT: Wagenaar, Melissa M.
; APPLICANT: Graziani, Edmund
; APPLICANT: Summers, Mia
; APPLICANT: Kulowski, Kerry
; APPLICANT: Pong, Kevin
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Production of a Complex
; TITLE OF INVENTION: Polyketide
; FILE REFERENCE: AM-101426US
; CURRENT APPLICATION NUMBER: US/11/143,980
; CURRENT FILING DATE: 2005-06-03
; PRIOR APPLICATION NUMBER: US 60/664,483
; PRIOR FILING DATE: 2005-03-23
; PRIOR APPLICATION NUMBER: US 60/576,895
; PRIOR FILING DATE: 2004-06-03
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
;   LENGTH: 116856
;   TYPE: DNA
;   ORGANISM: Streptomyces sp.
US-11-143-980-1

Query Match             2.5%;  Score 37.2;  DB 7;  Length 116856;
Best Local Similarity  61.2%;  Pred. No. 21;
Matches   60;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy   1398  CCCCGACGGCGCTGGGGGACAACTGGGCGTGGCGCTACCGCGAAGGCGACCTCGAGCCCGG 1457
              ||| ||||        ||||          ||||||     ||||
Db 105259  CCGGGGCCGGCTGAAGGACCAGCGGGGGTGGCCCGGGTGGCCCTGACTTGGCTGACCCGG 105318

Qy   1458  TCTGGCCGCGGGGACTGCGGGGCCCTGCGCCGGAGGCCAGCC 1495
              ||| |||||        ||||
Db 105319  GGTGACCCGGGCGGGGACCGGAGGGCGCCGCCGCGCCCCCAGGC 105356


RESULT 13
US-11-121-086-3
; Sequence 3, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
```

```
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 3
;   LENGTH: 168516
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-121-086-3

Query Match             2.5%;  Score 37.2;  DB 7;  Length 168516;
Best Local Similarity  46.4%;  Pred. No. 22;
Matches  156;  Conservative    0;  Mismatches  178;  Indels    2;  Gaps    1;

Qy   1110  TTCCGGTGAGGACAACGCCTTTTGCCCACAACTACCCGGCACGGCAATGTGGTGGT 1169
              |||||        |||||                |||||     ||||
Db 135213  TGCCGCTGACCTTCATGTCCATGGCCACGATCCTCCATGGCGCCAAGCGCCCTCGCGGATGG 135272

Qy   1170  GTACAGCGGAACCCACGACAACGACACACACCCTGGGATGGTTCCGCGCCGCCGGAGGC 1229
              |||||        ||||
Db 135273  ACACAGCCCCCGTGAGCACGCGCCCAGACCCGTGTGGGATGGGATGGAGCGCCACTCCGAGCGCGAGGC 135332

Qy   1230  CGAGCCGGCCTTCATGGGGCCTCACCTGGCCCGCTATGGCACCATTCCGTTGTTGTCGGAATA 1289
              |||||        |||||
Db 135333  AGAGCCAGTCGCAGAGAGGGCGGGGGTCGGCAGGAGCAGCAGCCCTGGCCCACTGAGCGCA 135392

Qy   1290  CGAGGTCGCGGGCGCTTTGATCGAGCTGGCCTTCAAAAGCCCGGCCAAGCTGGCTATTGT 1349
              |||||        |||||
Db 135393  GTTGGCAGCTGAGCCTGGCGGGGGACGCGGGCAGGAAGGGGCTGGGCTTCCCT 135452

Qy   1350  GCCTTTGCAGGACGTGCTGGGGCTGGGCCCCGA--GGCCCGCATGAACTTCCCCGGACACGG 1407
              |||||        |||||
Db 135453  GGGAGTGCTGGGCAGCCCGAGGCAGTGAAACGAGCGACCCACATTCAGCACCTCTGGGTG 135512

Qy   1408  CTGGGGGACAACTGGGCGTGGCGCTACCGCCGAAGGC 1443
              |||||        |||||
Db 135513  CTGGGGAAGCACACGTGCGCAGTGCCAGCCCTTGGGC 135548


RESULT 14
US-11-075-185-37
; Sequence 37, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
; APPLICANT: REEVES, CHRISTOPHER  D
; APPLICANT: JULIEN, BRYAN
; APPLICANT: REID, RALPH
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
; FILE REFERENCE: 010099.03
; CURRENT APPLICATION NUMBER: US/11/075,185
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/551,103
; PRIOR FILING DATE: 2004-03-08
; PRIOR APPLICATION NUMBER: US 60/568,290
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 37
;   LENGTH: 10524 
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-11-075-185-37

Query Match             2.4%;  Score 36.8;  DB 7;  Length 10524;
Best Local Similarity  52.6%;  Pred. No. 19;
Matches   80;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy   1330  CCGGCCAAGCTGGCTATTGTGCCTTTGCAGGACGTGCTGGGGCTGGGCCCCGAGGCCCGC 1389
              |||||        |||||
Db   4939  CCGGCGCCGCCGCTTCGCCGATCTCGCCCGCTCGAGCGAGCCGCAGCCGGCGCGCCGCTTCG 4998

Qy   1390  ATGAACTTCCCCGGACGGCGCTGGGGGACAACTGGGCGTGGCGCTACGCCGAAGGCGACCTC 1449
              |||||        |||||
Db   4999  CCGATGCTGGCGGCAGCTGGTGCGCGGCGCGCCGCGGCGGCGGCTACGCGGGGACGGCGGC 5058
```

```
Qy    1450  GAGCCCGGTCTGGCCGCGGGACTGCGGGCCCT  1481
            | |--|  ||||  ||||  |||  |||||  ||
Db    5059  GCGAGCGCGCAAGGCCTCGGCGCTGCGCGCGCT  5090


RESULT 15
US-11-128-061-384
; Sequence 384, Application US/11128061
; Publication No. US20060003958A1
; GENERAL INFORMATION:
;  APPLICANT: Melville, Mark W.
;  APPLICANT: Charlebois, Timothy S.
;  APPLICANT: Mounts, William M.
;  APPLICANT: Hann, Louane E.
;  APPLICANT: Sinacore, Martin S.
;  APPLICANT: Leonard, Mark W.
;  APPLICANT: Brown, Eugene L.
;  APPLICANT: Miller, Christopher P.
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS
;  TITLE OF INVENTION:  TO MONITOR GENE EXPRESSION
;  FILE REFERENCE: 01997.027701
;  CURRENT APPLICATION NUMBER: US/11/128,061
;  CURRENT FILING DATE: 2005-05-11
;  PRIOR APPLICATION NUMBER: US 60/570,425
;  PRIOR FILING DATE: 2004-05-11
;  NUMBER OF SEQ ID NOS: 7285
;  SOFTWARE: PatentIn version 3.3
;
; SEQ ID NO 384
;  LENGTH: 1987
;  TYPE: DNA
;  ORGANISM: Mesocricetus auratus
US-11-128-061-384

Query Match            2.4%;   Score 36.6;  DB 7;  Length 1987;
Best Local Similarity  50.3%;  Pred. No. 17;
Matches   90;  Conservative   0;  Mismatches   89;  Indels   0;  Gaps   0;

Qy    927  CCGGCCCAATGCTGTGGAGGGGGCGCTGGGTCAAAGCCCCAGGGAGAAGCTGTTTGCTGC  986
           ||||||  ||||  |||||| ||  ||||||| ||  ||||||||  |||  |||||||||
Db   1342  CCTGATGATTGGTCTCATCTGGCCGCTCCATGGAACACCCAGGGAAGCTCCTGTTTGCTCC  1401

Qy    987  GGTGCGGGCCCAACTGAGCGATGCGCCCATCATTGCCGAAGACCTGGGGGTGATCACCCC  1046
           ||  |||| |  |||||||||||| ||  |||||||||| ||  |||||||| ||  ||||
Db   1402  TAATTTGCTCCTGGACAGGAATCAGGGCAAGTGTGGAGGGCATGTGGGAGATCTTTGA  1461

Qy   1047  CGAGGTGGAGGCTTTGCGCGATGGCTTCGGGGTTCCCCGGCCATGAAGATTTTGCAGTTTG  1105
           ||||| |||| || ||  |  ||| | | |  ||  ||||  ||||  | |   |||||||
Db   1462  CATGTTGCTGCTGCTACATCAGCTCGGTTCCGCATGATGGACCTGCACGGGAGAGGAGTTTG  1520
```

Search completed: January 14, 2006, 17:24:45
Job time : 400 secs

```
                    GenCore version 5.1.6
         Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:      January 14, 2006, 06:57:40 ; Search time 880 Seconds
                                  (without alignments)
                                  11382.992 Million cell updates/sec

Title:       US-10-618-976-1
Perfect score: 1503
Sequence:    1 atgcaactccaacgcgcttt........ccgaggccagccagcgcgct 1503

Scoring table: IDENTITY_NUC
             Gapop 10.0 , Gapext 1.0

Searched:    496997 seqs, 333246308 residues

Total number of hits satisfying chosen parameters:    9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
             Maximum Match 100%
             Listing first 45 summaries

Database :   N_Geneseq_21:*
             1:  geneseqn1980s:*
             2:  geneseqn1990s:*
             3:  geneseqn2000s:*
             4:  geneseqn2001as:*
             5:  geneseqn2001bs:*
             6:  geneseqn2002as:*
             7:  geneseqn2002bs:*
             8:  geneseqn2003as:*
             9:  geneseqn2003bs:*
            10:  geneseqn2003cs:*
            11:  geneseqn2003ds:*
            12:  geneseqn2004as:*
            13:  geneseqn2004bs:*
            14:  geneseqn2005s:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1503	100.0	1503	5	AAD04867	Aad04867 Thermus r
2	638.4	42.5	1503	2	AAV72539	Aav72539 Thermus f
3	638.4	42.5	2085	12	ADQ15010	Adq15010 DNA encod
4	636.6	42.4	1503	10	ADC23306	Adc23306 DNA of Th
5	378	25.1	1539	13	ADU07728	Adu07728 DNA seque
6	329.2	21.9	1518	12	ADQ15025	Adq15025 DNA encod
7	289	19.2	1518	13	ADK45052	Adk45052 Streptoco
8	287.4	19.1	1515	10	ABX07913	Abx07913 S. pneumo
9	287.4	19.1	5910	2	AAV52298	Aav52298 Streptoco
c 10	287.4	19.1	110000	10	ABS56454_19	Continuation (20 o
11	265.6	17.7	1377	13	ADR91701	Adr91701 Novel S.
12	265.6	17.7	1377	14	AEA55571	Aea55571 Streptoco
13	181.4	12.1	1731	12	ADQ15021	Adq15021 DNA encod
14	173	11.5	1491	6	ABN69262	Abn69262 Streptoco
15	169.4	11.3	1494	6	ABN69261	Abn69261 Streptoco
16	169.4	11.3	1497	13	ADV84037	Adv84037 Streptoco
17	169.4	11.3	16857	13	ADV87699	Adv87699 Streptoco
18	169.4	11.3	16857	13	ADV78952	Adv78952 Streptoco
c 19	169.4	11.3	110000	6	ABN71527_14	Continuation (15 o
c 20	169.4	11.3	110000	13	ADV81204_15	Continuation (16 o
21	164	10.9	1581	12	ADQ15027	Adq15027 DNA encod
22	160.6	10.7	1458	13	ADC23308	Adc23308 DNA of Aq
23	159	10.6	1676	13	ADX60464	Adx60464 Plant ful
24	113.2	7.5	1485	14	ACL66091	Acl66091 M. xanthu
c 25	113.2	7.5	15377	14	ACL64653	Acl64653 M. xanthu
26	91.4	6.1	110000	2	AAX20248_07	Continuation (8 of
27	66.8	4.4	1945	13	ADX54712	Adx54712 Plant ful
28	64.6	4.3	259	7	ADS65998	Ads65998 Corn seed
29	64.2	4.3	1557	9	AAZ45687	Aaz45687 DNA encod
30	58.6	3.9	2097	6	ABA99452	Aba99452 Actinopla
31	58.6	3.9	38064	9	ABA99469	Aba99469 Actinopla
32	56.8	3.8	2217	6	ABZ12444	Abz12444 Arabidops
33	56.2	3.7	459	4	AAS07075	Aas07075 DNA encod
34	55.6	3.7	978	10	ABZ66744	Abz66744 Orthosomy
35	55.6	3.7	45055	10	ABZ66808	Abz66808 Orthosomy
c 36	54	3.6	2000	8	ADA71938	Ada71938 Rice gene
37	50	3.3	36401	14	ADY26621	Ady26621 Streptomy
38	49.4	3.3	1959	13	ADT17699	Adt17699 Plant cDN
39	49.4	3.3	2000	8	ADA71938	Ada71938 Rice gene
40	48.8	3.2	745	2	ADG89856	Adg89856 Corn 4-al
41	48.6	3.2	745	2	AAV70279	Aav70279 Corn 4-al
42	48.2	3.2	921	11	ACL33682	Acl33682 Rice abio
43	48.2	3.2	1146	8	ABT23235	Abt23235 Seed deve
44	47.6	3.2	2658	8	ACA37790	Aca37790 Prokaryot
c 45	47.4	3.2	1836	5	AAH67448	Aah67448 C glutami

ALIGNMENTS

```
RESULT 1
AAD04867
ID   AAD04867 standard; DNA; 1503 BP.
XX
AC   AAD04867;
XX
DT   06-AUG-2003  (revised)
DT   17-JUL-2001  (first entry)
XX
DE   Thermus rubens glucanotransferase gene.
XX
KW   Glucanotransferase; starch; Japanese dessert; noodles; wheat product;
KW   sea food; processed food; baby; animal; drink; gyoza skin; shumai skin;
KW   nutrient supplemental food; snack; detergent composition; ds.
XX
OS   Thermus sp.
XX
FH               Location/Qualifiers
FT   Key
FT   CDS          1..1503
FT                /*tag= a
FT                /product= "Thermus rubens glucanotransferase"
FT                /notes "Does not include stop codon"
FT                /partial
XX
PN   WO200129195-A1.
XX
PD   26-APR-2001.
XX
PF   16-OCT-2000; 2000WO-DK000560.
XX
PR   20-OCT-1999;  99DK-00001501.
PR   15-NOV-1999;  99DK-00001641.
XX
PA   (NOVO ) NOVOZYMES AS.
XX
PI   Fukuyama S;
XX
DR   WPI; 2001-290912/30.
DR   P-PSDB; AAB01153.
XX
PT   Novel polypeptides having glucanotransferase activity for producing food
PT   and in detergent or cleaning compositions for removing starch stains,
```

```
PT   especially amylose stains.
XX
PS   Claim 16; Page 49-52; 57pp; English.
XX
CC   The present DNA sequence is Thermus rubens ATCC 31556 glucanotransferase
CC   gene. Glucanotransferase catalyses the cyclisation of starch by acting on
CC   starch in food to produce a cyclic glucan. Glucanotransferase is useful
CC   for producing food such as Japanese desserts, snacks, wheat products,
CC   noodles, gyoza skins, shumai skins, processed sea foods, frozen or
CC   refrigerated processed foods, weaning foods, baby foods, pet foods,
CC   animal feeds, drink, sports foods and nutrient supplemental foods. A
CC   cleaning or detergent composition comprising glucanotransferase is useful
CC   for removal of starch stains, in particular for removal of amylose
CC   stains, from a hard surface or from laundry, by contacting the amylose
CC   stain-containing hard surface or the amylose stain-containing laundry
CC   with glucanotransferase or its composition. Nucleic acid constructs,
CC   vectors and host cells comprising glucanotransferase nucleic acid
CC   sequence are also provided. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ   Sequence 1503 BP; 248 A; 452 C; 497 G; 306 T; 0 U; 0 Other;

Query Match            100.0%;  Score 1503;  DB 5;  Length 1503;
Best Local Similarity  100.0%;  Pred. No. 0;
Matches 1503;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;
```

```
Qy      1 ATGCAACTCCAACGCGCTTTGGAATTTGCTCCACCCCACCAGTTTCCCGGTCCTGG     60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGCAACTCCAACGCGCTTTGGAATTTGCTCCACCCCACCAGTTTCCCGGTCCTGG     60

Qy     61 GGGATTGGGGCTCTGGGCCGCGAGGCCGAGCGGTTTTGGACTGGCTGGCCGATGCGGGGA  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GGGATTGGGGCTCTGGGCCGCGAGGCCGAGCGGTTTTGGACTGGCTGGCCGATGCGGGGA  120

Qy    121 GCCCGCTGGTGCAGTCTTACCGCTGGGCCCTACCAGTTACGGCGACTCGCCGTACCAG    180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCCCGCTGGTGCAGTCTTACCGCTGGGCCCTACCAGTTACGGCGACTCGCCGTACCAG    180

Qy    181 TCCTTCTCGGCTTTTGCGGTAACCCGTATTTGGTTGACCCGAGATGCTGATTGAAAAA    240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCCTTCTCGGCTTTTGCGGTAACCCGTATTTGGTTGACCCGAGATGCTGATTGAAAAA    240

Qy    241 GGCTGGCTGGAACAAAGCGAAGCGCCCCGTATCCGACCCAGCGCGTGGATTATGGC     300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGCTGGCTGGAACAAAGCGAAGCGCCCCGTATCCGACCCAGCGCGTGGATTATGGC     300

Qy    301 TGGCTTTACCAGACCCGCTGGCCCCTGTTCGGCGGGGTTTCGGCGGGTTTCGGCGGGAAGG  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TGGCTTTACCAGACCCGCTGGCCCCTGTTCGGCGGGGTTTCGGCGGGTTTCGGCGGGAAGG  360

Qy    361 GCTTCGGCCCAGGATAAGACCCGACTGGAAGCCTTTATCGAGGCGCCGAGCGCTTCTGGCTG  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCTTCGGCCCAGGATAAGACCCGACTGGAAGCCTTTATCGAGGCGCCGAGCGCTTCTGGCTG  420

Qy    421 GAAGACTATGGCTCTTTATGCCCTCTTCAAGACCCGGTTTGACGGCAAGCCCTGGAACGAG  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GAAGACTATGGCTCTTTATGCCCTCTTCAAGACCCGGTTTGACGGCAAGCCCTGGAACGAG  480

Qy    481 TGGAGCCCCGAGCTGCGCGACCGTGACCGTGAACCGGCTGCCCTGGCCCAGGGCCCGTGAGGAGCTG  540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TGGAGCCCCGAGCTGCGCGACCGTGACCGTGAACCGGCTGCCCTGGCCCAGGGCCCGTGAGGAGCTG  540

Qy    541 GCCGAGGAGGTGGCCCTTTACGAGGTGCCCCTTGTTCGGCTTTTTTTATCTGGAATGGGGGCCAG  600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GCCGAGGAGGTGGCCCTTTACGAGGTGCCCCTTGTTCGGCTTTTTTTATCTGGAATGGGGGCCAG  600

Qy    601 ACCAAGGCCTATGCGAATCCAAGGGGATTCAGATTATCGGCGATATGCCCATCTTTGTG  660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ACCAAGGCCTATGCGAATCCAAGGGGATTCAGATTATCGGCGATATGCCCATCTTTGTG  660

Qy    661 GCCTTCGATTCCTCAGATGTCTGGGCCAACCCGCAGTACTTCTACCTGCAGCCGATGGC  720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GCCTTCGATTCCTCAGATGTCTGGGCCAACCCGCAGTACTTCTACCTGCAGCCGATGGC  720
```

```
Qy    721 AACCCCACGGTGGTGCGCGGGCGTTCCGCCGGGACTACTTCTCCGAAACCGGCCAGCTCGG  780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AACCCCACGGTGGTGCGCGGGCGTTCCGCCGGGACTACTTCTCCGAAACCGGCCAGCTCGG  780

Qy    781 GGCAATCCGCTCTATCGCTGGGATGGATGATGAAAGGGACAAACTTGCCTGGTGCATTGCC  840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GGCAATCCGCTCTATCGCTGGGATGGATGATGAAAGGGACAAACTTGCCTGGTGCATTGCC  840

Qy    841 CGCATAAGGCAGTCGCTCAAGCAGTGCCACCTGGTGCCATCGACCAGTTCCGCGGGGTT   900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CGCATAAGGCAGTCGCTCAAGCAGTGCCACCTGGTGCCATCGACCAGTTCCGCGGGTTT   900

Qy    901 GAAGGCCTACTGGGAGGTTCCGTTGCCGGCCAATGCTGTGGAGGGCGCTGGGGTCAAA    960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GAAGGCCTACTGGGAGGTTCCGTTGCCGGCCAATGCTGTGGAGGGCGCTGGGGTCAAA    960

Qy    961 GCCCCAGGGGAGAAGCTGTTTGCTGCGGTGCGGGCCCAACTGAGCGATGCGCCCATCATT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GCCCCAGGGGAGAAGCTGTTTGCTGCGGTGCGGGCCCAACTGAGCGATGCGCCCATCATT 1020

Qy   1021 GCCGAAGACCTGGGGGTGATCACCCCCGAGGTGGAGCGTTGCGCGATGGCTTCGGGTTC  1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GCCGAAGACCTGGGGGTGATCACCCCCGAGGTGGAGCGTTGCGCGATGGCTTCGGGTTC  1080

Qy   1081 CCCGGCATGAAGATTTTGCAGTTTGCTTTTTCCGGTGAGGACAACGCCCTTTTTGCCCCAC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CCCGGCATGAAGATTTTGCAGTTTGCTTTTTCCGGTGAGGACAACGCCCTTTTTGCCCCAC 1140

Qy   1141 AACTACCCCGCGCACGGCAATGTGTGGTACAGCGGACCCAGCACAACGACACACCACCACC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 AACTACCCCGCGCACGGCAATGTGTGGTACAGCGGACCCAGCACAACGACACACCACCACC 1200

Qy   1201 CTGGGATGGTTCCGCACCGGCCGGCCGAGCGGCCCTTCATGCGGCCTTCATGCGGGCTCACCTGCC 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CTGGGATGGTTCCGCACCGGCCGGCCGAGCGGCCCTTCATGCGGCCTTCATGCGGGCTCACCTGCC 1260

Qy   1261 CGCTATGGCATCCGTGTTTGTTGGGAATACAGAGTCGCGGGCGCTTTGATCGAGCTCGCC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CGCTATGGCATCCGTGTTTGTTGGGAATACAGAGTCGCGGGCGCTTTGATCGAGCTCGCC 1320

Qy   1321 TTCAAAAGCCCGGCCAATGAACTTCCCCGGACGCTGGGGGGACAACTGGGCGTCGCCGAA  1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TTCAAAAGCCCGGCCAATGAACTTCCCCGGACGCTGGGGGGACAACTGGGCGTCGCCGAA  1380

Qy   1381 GAGGGCCGGCCATGAACTTCCCCGGACGCTGGGGGGACTGGGCCCTGCGCGAGGCCCAGCGC 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 GAGGGCCGGCCATGAACTTCCCCGGACGCTGGGGGGACTGGGCCCTGCGCGAGGCCCAGCGC 1440

Qy   1441 GGCGACCTGAGCCCCGTCTGCGCGGGCCCTGCGCGAGCCCGCGGGGACTGCGCCAGCCAGCGC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GGCGACCTGAGCCCCGTCTGCGCGGGCCCTGCGCGAGCCCGCGGGGACTGCGCCAGCCAGCGC 1500

Qy   1501 GCT  1503
          |||
Db   1501 GCT  1503
```

```
RESULT 2
AAV72539
ID   AAV72539 standard; DNA; 1503 BP.
XX
AC   AAV72539;
XX
DT   17-OCT-2003 (revised)
DT   01-MAR-1999 (first entry)
XX
DE   Thermus flavus amylomaltase encoding DNA.
XX
KW   Thermus flavus; amylomaltase; heat resistant; cyclic glucan;
KW   intramolecular transglycosylation; alpha-glucan; food; ds.
XX
OS   Thermus thermophilus.
XX
```

```
FH   Key             Location/Qualifiers
FT   CDS             1..1503
FT                   /*tag= a
FT                   /product= "amylomaltase"
XX
PN   EP884384-A2.
XX
PD   16-DEC-1998.
XX
PF   13-MAY-1998;   98EP-00250162.
XX
PR   13-MAY-1997;   97JP-00122635.
PR   07-MAY-1998;   98JP-0125121.
XX
PA   (EZAK ) EZAKI GLICO CO LTD.
XX
PI   Terada Y, Fujii K, Yanase M, Takata H, Takaha T, Okada S;
XX
DR   WPI; 1999-026580/03.
DR   P-PSDB; AAW83330.
XX
PT   New amylo-maltase catalyses intra-molecular trans-glycosylation of alpha-
PT   glucans - used as additive to e.g. rice products, snacks, wheat products,
PT   noodles, processed seafood, frozen or refrigerated foods, baby foods or
PT   drinks.
XX
PS   Claim 5; Page 19; 32pp; English.
XX
CC   The present sequence encodes amylomaltase isolated from Thermus flavus.
CC   Amylomaltase catalyses intramolecular transglycosylation of alpha-glucans
CC   to generate cyclic glucans, has no hydrolase activity, has an optimum
CC   temperature of 65-70 degrees Celsius, remains active at 60 degrees
CC   Celsius for at least 10 minutes, is inactivated at 100 degrees Celsius
CC   within 15 minutes and has an optimum pH of 5.5. Amylomaltase can be used
CC   to produce a cyclic glucan by cyclising an alpha-glucan and collecting
CC   and purifying the cyclic glucan (especially where the cyclic glucan
CC   comprises a cyclic alpha-1,4-glucan, a branched cyclic glucan and/or a
CC   branching enzyme is also used in the cyclisation step). The amylomaltase
CC   is used to produce food by adding it to a food material before or
CC   immediately after cooking so that the amylomaltase acts on starch in the
CC   food material to produce a cyclic glucan (especially where the food is
CC   selected from rice products, Japanese desserts, snacks, wheat products,
CC   noodles, gyoza skins, shumai skins, processed seafoods, frozen or
CC   refrigerated processed foods, weaning foods, baby foods, pet foods,
CC   animal feeds, drinks, sports foods and nutritional supplements). (Updated
CC   on 17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1503 BP; 223 A; 517 C; 523 G; 240 T; 0 U; 0 Other;

Query Match          42.5%;   Score 638.4;  DB 2;  Length 1503;
Best Local Similarity 65.0%;   Pred. No. 1.9e-142;
Matches 977;  Conservative 0;  Mismatches 516;  Indels 9;  Gaps 2;
```

```
Qy    1 ATGCAACTCCAACGCGCTTTTGGAATTTGCTCCACCCCACCAGTTTCCGGGGTCGCTGG 60
          ||||||    ||||||||||  ||| || |||||||||||||        | ||||| |||
Db    1 ATGGAGCTTCCCCCGCGCTTCCGGTCTCGTCTTCCACCCCACGAGCCTCCCCGGCCCCTAC 60

Qy   61 GGGATTGGGCGGTCTCGGCCGCGGCAGGCCGAGCGGTTTTGGACTGGACTGGCGATGCGGGA 120
          ||||| || | || ||||||||| |||| ||||| |||| |||||| |||  ||||||||| 
Db   61 GGGCGTCGGGCGTCCTGGGCGGAGGCCCGGGGACTTCCCGCTTCCTCCGGGGACTTCCTCC 120

Qy  121 GCCCCGTCGTGGCAGGTCTTACCGCGTGGGCGCCCTACCAGTTACGGCGACTCGCCGTACCAG 180
          |||||| | || |||  || |||||| || ||||  ||||||||| ||||||||  ||||||| 
Db  121 GCCGGGTACTGGCAGGTCTCTCCCCCTTGGGGCCCTATGGGGCCGTATGGCCGCGTGTTACCAG 180

Qy  181 TCCTTCTCGGCTTTTGCCGTAACCCGTATTTGGTTGACCCGACGATGCGTGATTGAAAA 240
          ||||| || || || ||||| | ||| ||||||||| ||| |||||||| |||||||| 
Db  181 TCCTTCAGCGGCTTTTCCGGGAAACCCTACCTCTCATGAGACCTGAGGCCCTCTCGGGGAAAGG 240

Qy  241 GGCTGGCTGGAACAAAGCGAAGGCGCCCCCGCCCGGCTTCCCCAAGGCCCGCGTGTGGATTATGGC 300
          ||||          | |||| |||||||||| |||||| ||||||| || ||||| ||| || ||| 
Db  241 GGCTACGTG------CGCTTGGAGGACCCCGGCTTCCCCAAGGCGGTGGGGACACTGGGCGTCGTACGCC 294

Qy  301 TGGCTTTACCGACCCGCTGGCCCTGTGCCGGGGCGTTTCGGCGGGTTCGGGGCAAGG 360
          ||    |||| | || || ||||||| ||| |||||||||| ||||| ||  |||| 
Db  295 CTCCTCTACGCCTGGACTGGCCTCGGAAGCCTTGAGAGGAGGCCTTCAAGGAAAAG 354

Qy  361 GCTTCGGCCCCAGGATAAGACCCGACTGGAGCCGCCTTATCGCCGGGAGGCGCTTCGGGCTG 420
          | |||| | ||| || | ||| |||  || ||||||| || |||||| | |||||| || |||
Db  355 GCCTCCCCGGAGGAGGAGGAGGAGGCCTTCGCCGCCTTCCCCGGAGAGGAGGCCTGGTGGCTC 414

Qy  421 GAAGACTATGCGCTCTTTTATGGCCCTCAAGACCCGGTTTGACCGCAAGCCCTGAAACGAG 480
          || ||| ||||| | ||| ||||||||||| | | | ||||| |||||| || ||||||| 
Db  415 GAGGACTACGCCCTCTTCATGGCCCTGAAGGGCCACGGGGGGCGCACGGGGGGCTTCCCTGGAACCGG 474

Qy  481 TGGAGCCCCGAGTGCGCGACCGTGAACCGGGTCCCTGCCCAGGGCCCGTGAGGAGCTG 540
          |||            || |||| | | ||||| || || |||||||| ||| ||||| 
Db  475 TGGCCCCTTCCCCTGCCCCCTGCGGGAAGCCGGAGAAGAAGCCCCGTTAGGGAGGCGAAAAAGCGCCTTG 534

Qy  541 GCCGAGGAGGTGGCCCTTTACGAGTGGATTCAGTGGCTTTTTATCTGGAATGGGGCCAG 600
          ||||||||||||| ||||  |||  || ||||| || | |||  | ||| |||||| ||
Db  535 GCCGAGGAGGTGGCCTTCACCCAGGGCTCTTCTTCACCCAGGGCTCTTCCGGCCAGTGGGGGGGGC 594

Qy  601 ACCAAGGCCTATGCCGAATCCAAGGGGATTCAGATTATCGGCGATGATGATGCCCATCTTGTG 660
          | |||  || ||||| ||||| ||| | |||  || |||||||  || |||||||||||||| 
Db  595 TTGAAGGCGGAGGCCGAGGCGGTGCCCCCACTACTTCTCGGAGAAAGGGAACATGCCCATCTTCGTG 654

Qy  661 GCCTTCGATTCTCGAGATGTCTGGCCCAACCCGCAGTACTTTCACCTGGACGACGAGGGC 720
          | ||| |||| | |||| || |||  | ||||| |||||  ||| ||| |||| || ||||||
Db  655 GCCGAGGACTCCGGCGAGTCTGGGGCGTCTGGGGCCCACCCCGAGTGGTTTCACCTGGACGACGAGGGC 714

Qy  721 AACCCCACGGTGGTGCGGCGTTCGCGGGGCGGTTCGGCCAGCTCCTG 780
          |||||| | |||||| | ||||| || || |  ||||||||||| || 
Db  715 CGCCCCCACGGTGGTGCGGGGGGTGCCCCCGACTACTTCCGGGGGCCCAGCGCCCAGCGCCTG 774

Qy  781 GGCAATCCGCTCTATCGCTGGGATGTGATGATGGAAAGGGACAACTTGCTGGTGCATTGCC 840
          | |||||| ||| |||||||||| | || ||||| |  ||| | || ||| |||||| ||
Db  775 GGTAACCCCTTTACCGCTGGGACGTTTTGGAGCGGGGAGGGGTTCTCCTTCTCCTTCTGGATCCCC 834

Qy  841 CGCATAAGGCAGTCGCTCAAGCAGTGCCACCTGGTGCGCACCACTTCCCGGGGTT 900
          ||| | || || || ||  | |||| ||| | |||||| ||||| |||||  ||| 
Db  835 CGTCTGGAGAAGGCCCTGGAGCTCTTCACCTGGTGCGCCATAGACCACTTCCGGCGGCTTT 894

Qy  901 GAAGCTACTGGGAGGTTCCGTTTGGCCCCAATCCTGTGGAGGGGCGCTGGGTCAAA 960
          |||||| |||||| || || |||  | ||| |||| ||| |||| |||||| |||
Db  895 GAGGGCCCTACTGGGAGATCCCCGGCCCAGCTGCCCCCACGGGGGCGGTGAGGGGGCGCTGGGTCAGG 954

Qy  961 GCCCCAGGGGACGAAGCTGTTGCTGGCGTGGGGCCCAACTGAGCGATGGCGCCCATCATT 1020
          |||||  ||| || || || ||||| |  ||| | | |||   ||||| |||| ||||| 
Db  955 GCCCCGGGGGGGAGAAGCTCTTCGGCAGATCCAGAAGATCCAGAAGATCCAGCTTCTTCGGCGGATGGCGCCGGAGGTCTCCCGGTCTC 1014

Qy 1021 GCCGAAGACCTGGGGGTGATCACCCCGAGGTGGAGGCTTTGCGGCGATGGCTTCGGGGTC 1080
          ||||| ||| ||||| |||||| ||||| |||| |||| ||| || |||| | |||||| 
Db 1015 GCCGAGGAGCTGGGGGTCATCACCCCGAGGTGGAGCCCGCGGGAGCGGGACCGGCTTCGGCCGCCTTT 1074

Qy 1081 CCCGGCCATGAGAGATTTTGCAGTTTGCTTTT---TCCGGTGAGGACAACGCGCCTTTTGCCC 1137
          ||||| |||||| ||| || ||| |||||||   ||| || ||||| ||  ||||| || ||
Db 1075 CCCGGGGATGAGAGGTCCTGCAGTTGCCCTTTGACGACGGGATGAAAAACCCCTTCCTCCCCC 1134

Qy 1138 CACAACTACCCCGCGCAATGTGGTGTACAGCGGAACCCACGACAACGACACC 1197
          |||||||||||| |||    || ||||  |||||||| |||||| ||||||| 
Db 1135 CACAACTACCCTGCCCACCGGTGGTGTCTACACCGGCGGCCTTCATGGGGGCCTTCATGTGGGCCGAGGGGCCTTCATGAGGGGCTACCTG 1194

Qy 1198 ACCCTGGATGTTCCGCAGGTCTTTACCGCGTTACGGCCCTACCAGTTACGGCGACTCGCCGTACCAG 1257
          |||||||| || || || ||||| ||| | ||||||  |||||||| ||||||||| |||  ||||| 
Db 1195 ACCCTGGGCTGGTACCGCACGGCCACCCCCACGGGCCCTTCATGAGGGGCGGTGGGTACCTG 1254

Qy 1258 GCCCGCTATGGCATCCGTTGTTGTTCGGAATACGAGCTCGGGGCGCTTTGATCGAGCTG 1317
          ||||||| || || ||||| ||| || |||| || ||||| |||| |||| || |||| 
Db 1255 GCGGGGCTGGGGGATCACCTTCGGGGGATCACTTCGGGGGCGCCCTGGGGCGCCGTGCACCTG 1314

Qy 1318 GCCTTCAAAAGCCCGGCCAAGCTGCCTTTTGCCTATTGTGCCTTTGCTTGGGCGTGCGGGC 1377
          ||||||||   || || |||  |  | ||||| || || |||||| |  ||||| ||||| 
Db 1315 GGGATGAAGTCCGTGGGCCCGGCCCTCTACCCGGTGGATTATGGC 1374

Qy 1378 CCCGAGCGCCCGCCATGAACGAACTTCCCCCGAAGGCCGGGACACTGGGCGTCGTACGCC 1437.
          |||||||||||| |||  | ||||| ||||| ||||| | ||||||||||||||| ||||| 
Db          ...
```

```
Db   1375 AGCGGAGGCCCGGATGAACTACCCGGGAAGGCCCTCGGGGAACTGGGCCTGGCCGCTCCTC 1434
          |||||||| ||||||| |||  |||||||||||||| |||| ||||||   ||||||
Qy   1438 GAAGGCGACCTCGAGCCCGGTCGCCGGCGGGGACTGCGGGGCCCTGGCCGAGGGCCAGCCAG 1497
          || || |  ||||||||| ||| ||| ||||| |  |||| |||||||||||||||||||
Db   1435 CCGGGGGAGCTTCCCCGAGCACGGGGCGCAGGCTTAGGGCCCATGGCCGAGGGCCACGGAA 1494

Qy   1498 CG 1499
          ||
Db   1495 CG 1496
```

RESULT 3
ADQ15010
ID   ADQ15010 standard; DNA; 2085 BP.
XX
AC   ADQ15010;
XX
DT   23-SEP-2004 (first entry)
XX
DE   DNA encoding Thermus aquaticus amylo maltase polypeptide.
XX
KW   Amylo maltase; transgenic; foodstuff; food additive; modifier;
KW   saccharide; (alpha)-1,4-glucan; cooked rice; Japanese confectionery;
KW   snack; noodle; Chinese dumpling; shao-mai; fishery kneading;
KW   refrigerated; baby food; pet; animal feed; drink; food supplement;
KW   cyclic glucan; Thermus aquaticus; gene; ds.
XX
OS   Thermus aquaticus.
XX
FH   Key   Location/Qualifiers
FT   CDS   514..2016
FT   /tag= a
FT   /EC_number= "EC 2.4.1.25"
FT   /product= "Thermus aquaticus amylo maltase"
XX
PN   JP2004187674-A.
XX
PD   08-JUL-2004.
XX
PF   12-SEP-2003; 2003JP-00322319.
XX
PR   26-NOV-2002; 2002JP-00342966.
XX
PA   (EZAK) EZAKI GLICO CO LTD.
PA   (NIDE) NEC CORP.
XX
DR   WPI; 2004-528719/51.
DR   P-PSDB; ADQ15011.
XX
PT   New mutated amylo maltase polypeptide having increased enzyme activity
PT   and reduced hydrolyzing activity, useful for manufacturing foodstuffs
PT   such as Japanese confectionery, noodles, baby foods, and food additives.
XX
PS   Claim 5; SEQ ID NO 1; 5lpp; Japanese.
XX
CC   The invention relates to a novel amylo maltase polypeptide comprising the
CC   amino acid sequence of a wild-type amylo maltase with a substitution,
CC   addition or deletion at a position, where the amino acid residue
CC   interacts with acarbose of amino acids other than the wild-type amylo
CC   maltase amino acid sequence. The invention further comprises: a nucleic
CC   acid molecule containing a nucleic acid sequence encoding the novel amylo
CC   maltase polypeptide; a vector containing the nucleic acid sequence; a
CC   cell containing the nucleic acid sequence; biological tissue containing
CC   the nucleic acid sequence; a transgenic organism containing the nucleic
CC   acid sequence; a foodstuff, a food additive or a modifier of foodstuff
CC   containing the nucleic acid sequence; a computer readable recording
CC   medium which contains the information of the nucleic acid sequence
CC   encoding the amino acid sequence of the novel amylo maltase polypeptide;
CC   and a cyclic glucan obtained by reacting the novel amylo maltase
CC   polypeptide on saccharides which have a linear structure of (alpha)-1,4-
CC   glucan. The novel amylo maltase polypeptide is useful for manufacturing
CC   foodstuffs such as cooked rice, Japanese confectionery, snacks, bakeries,
CC   noodles, Chinese dumpling, shao-mai, fishery kneading goods,
CC   refrigerating foods, baby foods, pet foods, feed for animals, drinks, and
CC   food supplements, where the method involves adding the novel amylo
CC   maltase polypeptide to the foodstuff before or immediately after heat
CC   processing of the raw material, where the novel amylo maltase polypeptide
CC   generates a cyclic glucan from the starch of the foodstuff. The novel
CC   amylo maltase polypeptide is useful for manufacturing glucan which has a
CC   cyclic structure of (alpha)-1,4-glycoside linkage, foodstuff materials
CC   and food additives, modifier of foodstuffs, a drink or eatable
CC   composition, infusion solution or composition for adhesion, where the
CC   method involves reacting the saccharide containing the linear structure
CC   of (alpha)-1,4-glycoside linkage or its derivatives with the novel amylo
CC   maltase polypeptide. This polynucleotide sequence represents the DNA
CC   encoding the Thermus aquaticus amylo maltase polypeptide of the
CC   invention.
XX
SQ   Sequence 2085 BP; 318 A; 715 C; 724 G; 328 T; 0 U; 0 Other;
XX

Query Match   42.5%; Score 638.4; DB 12; Length 2085;
Best Local Similarity   65.0%; Pred. No. 2.1e-142;
Matches 977; Conservative   0; Mismatches 516; Indels 9; Gaps 2;

```
Qy      1 ATGGAACTCCAACGCGCTTTTGGAATTTGCTCACCCCACCAGTTTCCGGGTCGCTGG 60
          ||||  ||  |  ||||| |  |||| ||| |  ||||| |||||||||||||||
Db    514 ATGGAGCTTCCCCGGCTTTGGTCTTCTCCACCCCACCCACGAGCCTCCCGGGCCCCTAC 573

Qy     61 GGGATTGGGGCTCTGGGCCCGAGGCCGAGCGGTTTTGACTGGCTGGCCGATGCGGGGA 120
          ||| |||||| || |||||||| || || ||||   |  | | |||||||||||||||
Db    574 GGCGTCGGCGTCCTCGGCCCGGCGGAGGCCCGGGACTTCCTCCGCTTCCTCAAGGAGGCGGGGG 633

Qy    121 GCCCGCTGGTGGCAGGTCTTACCGCTGGGCCCTACCGGCCGACTCCGCCGTACCAG 180
          |  ||||| |||| |  || | ||  |||||||| |||||||| ||| ||| ||||
Db    634 GGGCCGGTACTGGCCAGGTCCTCCCCCTCGGGCCGCATGGGACTCCCCCCTACCAG 693

Qy    181 TCCTTCTCGGCTTTTGCCGGTAACCCGTATTTGGTTGACCCCGAGATGCTGATTGAAAAA 240
          ||| ||  | |  |||||||  ||||| ||  |||| ||||  ||||| | ||| ||
Db    694 TCCTTCAGCGGCTTCGCCGGAAACCCTACCTCATAGACCTGAGGCCCTCGCGGAAGG 753

Qy    241 GGCTGCGCTGGAACAAAGCGAAGCGCCCCCGCCTATCGACCCAGGCGCGTGGATTATGGC 300
          ||   |||    || |  | |||||| ||||||  |||||||| |||||||||||||||
Db    754 GGCTACGTG------CGCCTGGAGACCCGCGCCTTCCCCAAGGGCCGGGTGGACTACGGC 807

Qy    301 TGGCTTTACCAGACCCGCTGCCCCGTGTTGCGGCGGGGCTTCGCGGCGGGGTTTCGGGCAAGG 360
          ||  ||||||| || ||  | |||  ||| ||||| || |||||||||| |||||||||||
Db    808 CTCCTCTACGCCTGGAAGTGGCCCGCCGGGGAGGAGGCCGTTCCGGGGCTTCAAGGAAAAG 867

Qy    361 GCTTCGGGCCCAGGATAAGACCCGACTGGAAGCCTTTATCGAGGCCGCGCTTCTGCCTG 420
          ||  ||||  |||| | |  |  |  ||||| ||| || |||||||| ||  |||||  |
Db    868 GCCTCCCCGGAGGAGGCGGAGGCCTTCGCCGCCTTCCGGGGAGGAGGGAGGCCTGGTGGCTC 927

Qy    421 GAAGACTATGCGCTCTTTATGCGCCTCAAGACCCGGTTTGACGCGAAGCCTGGAACGAG 480
          ||||| |||||  | ||  |||  ||| |||||| ||| ||  | |||  ||||| ||||
Db    928 GAGGACTACGCCCTCTTCATGGCCGTGAAGGGGCACGGGGGCGGCTTCCCTGGAACCGG 987

Qy    481 TGGAGCCCGAGCTGCGCGACCGTGAACCCGGGCTGCCCTGGCAGGGCCCGTGAGGAGCTG 540
          ||||||  | |||| ||||||| |||||||| ||| ||  ||| ||||  | ||||| ||
Db    988 TGGGCCCCTTCCCCTGCGGAAGCGGGAAGAAGAGGCCCTTAGGGAGGCGAAAAGCGCCTTG 1047

Qy    541 GCCGAGGAGGTGGCCCCTTTACGAGTGGATTCAGTGGCTTTTTTATCTGGAATTGGGGCCAG 600
          ||||| || | |||  |  |||| |  |||| |||||||  |  |  |||  | ||||||||
Db   1048 GCCGAGGAGGTGGCCTTCACCGTGGTCTTCTCCCCAGTGGGGGGCC 1107

Qy    601 ACCAAGGCCTATGCCGAATCCAAGGGATTCAGATTATCGGGATATGCGGCGATATGCGGGTTGTG 660
          ||  |||| | |||   ||  | |||  |  |||    |  ||   ||  ||||   |  |||
Db   1108 TTGAAGGCCGAGGCCGAGGCGGTTGGCCATCCGGATCATCCGGGACCATGCCCATCTTCGTG 1167

Qy    661 GCCTTCGATTCCTCAGATGTCTGGGCCCAACCCGACTACTTCACCTCGGCCGATGGC 720
          ||  ||| |  |  || |||| ||||| | |||| |||| |  |||  | || |||||
Db   1168 GCCGAGGACTCCGCCGAGGTCTGGGCCCACCCCGAGGTGGTTTCACCTGGACGAGGAGGGC 1227

Qy    721 AACCCCACGGTGGTGCGCGGCGGCGTTCCGCGGGACTACTTCTCCGAAACCGGCCAGCTCGG 780
          ||||| |||||||| | |||||||| | || |  || |||  ||||  || ||||||| |||
Db   1228 CGCCCCACGGTGGTGCGGGGGGGTGCCCCCGACTACTTCTCCGAGACGGGGCCAGCGCGTGG 1287
```

XX
PD   09-JAN-2003.
XX
PF   01-JUL-2002; 2002WO-NL000427.
XX
PR   29-JUN-2001; 2001NL-01018426.
XX
PA   (NEDE) NEDERLANDSE ORG TOEGEPAST.
XX
PI   Van Der Maarel MEC,   Dijkhuizen L,   Binnema DJ,   Van Der Veen BA;
PI   Vos A;
XX
DR   WPI; 2003-201501/19.
DR   P-PSDB; ADC23307.
XX
PT   Novel isolated or recombinant nucleic acid encoding a polypeptide with
PT   alpha glucanotransferase activity, useful for reducing the retrogradation
PT   of starch, amylopectin, and for hydrolyzing starch.
XX
PS   Disclosure; Page 78-79; 92pp; English.
XX
CC   The invention relates to a novel isolated or recombinant nucleic acid
CC   derived from a nucleic acid encoding a polypeptide essentially having
CC   alpha-glucanotransferase activity but having essentially no hydrolysing
CC   activity, and the recombinant nucleic acid encoding a polypeptide with
CC   hydrolytic activity. The invention further comprises a vector comprising
CC   the recombinant nucleic acid; a host cell comprising the vector or the
CC   recombinant nucleic acid. The invention also provides a polypeptide or
CC   its fragment essentially having alpha-glucanotransferase activity and
CC   with or without hydrolysing activity, by providing a nucleic acid
CC   encoding such a transferase with a mutation leading to an alteration or
CC   loss of a codon originally encoding a hydrophobic amino acid located in
CC   or around an acceptor, a donor or a catalytic site extending from a
CC   (alphalpha/betabeta) barrel structure of the transferase. The invention
CC   also relates to a bakery ingredient comprising the polypeptide with alpha
CC   -glucanotransferase activity and a bakery product such as bread,
CC   comprising the polypeptide. The transferase encoded by the recombinant
CC   nucleic acid comprises amylomaltase, a branching enzyme or thermostable
CC   transferase. The transferase is derived from a thermophilic microorganism
CC   e.g. Thermus thermophilus, T.aquaticus or Aquifex aeolicus. The
CC   transferase with enzyme commission (EC) number 2.4.1.25 or 2.4.1.18 is
CC   useful in reducing retrogradation of starch and amylopectin, preferably
CC   in reducing long-term retrogradation of amylopectin. This transferase is
CC   also useful for hydrolysing starch. This polynucleotide sequence
CC   represents the DNA encoding a Thermus thermophilus AMase protein of the
CC   invention.
XX
SQ   Sequence 1503 BP; 225 A; 518 C; 521 G; 239 T; 0 U; 0 Other;

Query Match   42.4%; Score 636.8; DB 10; Length 1503;
Best Local Similarity   65.0%; Pred. No. 4.7e-142;
Matches 976; Conservative 0; Mismatches 517; Indels   9; Gaps   2;

Qy   1 ATGCAACTCCAACGCGCTTTTGGAATTTGCTCACCCACCAGTTTTCCGGGTCGCTGG 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGAGCTTCCCCGCGCTTTCGGTCGTTCGGGTCTTGCTTCTCCACCCCAGAGCCTCCCCGCCCCTAC 60

Qy   61 GGGATTGGGGCTCTGGGCGCCGAGGCCGAGGCCCGGAGACTCCGATGCGGGA 120
          ||||||||||| ||||||| ||||||||
Db   61 GGGCTCGGCGTCCTGGGCCCAGGAGCCTCGGAGGCCCGCTTCCTCAAGGAGGCAGGG 120

Qy   121 GCCCGCTGGTGGCAGTCTTACCCGTGGGCCCTGGGCCCTACCAGTTACGGCGTCGCCGGACTCGCCGTACCAG 180
          |||||||||||||||||| ||||| ||| ||||||||||||||||| ||||||||||
Db   121 GGGCGGTACTGGCAGGTCCTCCCCTTGGGCCCCACGGGCTATGGGCGACTCCCCTACCAG 180

Qy   181 TCCTTCTCGGCTTTTGCCGGTAACCCGTATTTGGTTGACCCCGAGATGCTGATTGAAAAA 240
          ||||| |||||||||| ||||||||||| |||||||||
Db   181 TCCTTCAGGCGCTTCGCCGGAAACCCTACCTCATAGACTCGAGGCCCTCGAGGGCGAAAGG 240

Qy   241 GGCTGGCTGGAACAAAGGAGAGGCGCCCCCGCGTATCGCACCCGCGCGTGATTATGGC 300
          ||||| ||||| | ||||||||| |||||||||
Db   241 GGCTACGTG------CGCCTGGAGGACCCCCAAGGCCTCCCCAAGGGCCGGGTGGACTACGGC 294

RESULT 4
ADC23306
ID   ADC23306 standard; DNA; 1503 BP.
XX
AC   ADC23306;
XX
DT   18-DEC-2003   (first entry)
XX
DE   DNA of Thermus thermophilus AMase.
XX
KW   recombinant; alpha-glucanotransferase activity; hydrolysing; hydrolytic;
KW   alphalpha/betabeta; barrel structure; transferase; bakery product; bread;
KW   amylomaltase; branching enzyme; thermostable transferase;
KW   thermophilic microorganism; amylopectin; EC 2.4.1.25; EC 2.4.1.18;
KW   starch retrogradation; amylopectin; alpha-amylase; AMase; gene; ds.
XX
OS   Thermus thermophilus.
XX
PN   WO200302728-A2.

Qy   781 GGCAATCCGCTCTATCGCTGGGATGTGATGAAAAGGGACAACTTTGCCTGGTGCATTGCC 840
          ||| ||||||||||||||||
Db   1288 GGTAACCCCCTTTACCGCTGGGACGTTTTGGAGCGGGAGAGGGGTTCTCCTTCTCGGATCCGC 1347

Qy   841 CGCATAAGGCAGTCGCTCAAGCAGTGCCACCTGCTGCCATCGACCACTTCCGCGGGTTT 900
          |||||| ||||||| | ||||| ||||||||
Db   1348 CGTCTGGAGAAGGCCCTGGAGCTCTTCCACCTGGTGCCCATAGACCACTTCCGCGGGCTTT 1407

Qy   901 GAAGCCTACTGGGAGGTTCCGTTTGGCCCGGCCCAATGCTGTGGAGGGGCGCTGGGTCAAA 960
          ||||||||||||| || |||| |||||
Db   1408 GAGGCCTACTGGGAGATCCCCAAGCTGCCCCACGGGGTGGAGGGGCGCTGGGGTCAAG 1467

Qy   961 GCCCCAGGGAGGGGAAGCTGTTTGTCGGCGTGCGGCCCAACTGAGCGATGCGCCCATCATT 1020
          |||||||||||||||||||||
Db   1468 GCCCCGGGGAGGGGAGAGAGCTCTTCCAGAAGATCCAGGAGGTCTTCGGCGAGTCCCCGTCCTC 1527

Qy   1021 GCCGAAGACCTGGGGTGATCACCCCCGAGGTGGAGGCTTGGCCGATGGCTTCGGGTTC 1080
          ||||||||||||||||||||| ||
Db   1528 GCCGAGGACCTCGGGGTCATCACCCCCGAGGTGGAGGCCCTGCGACCGACCGCTTCGGGCCTT 1587

Qy   1081 CCCGGCATGAAGAGATTTGCAGTTTGCTTTT---TCCGGTGAGGACAACGCCCTTTTGCCC 1137
          |||||| ||||||||
Db   1588 CCCGGGATGAAGGTCCTGCAGTTCGCCCTTTGACGACGGGATGGAAAAACCCCTTCCTCCCCC 1647

Qy   1138 CACAACTACCCCGCGCACGGCAATGTGGTGTACAGCGAACCCACGACAACGACACCC 1197
          ||||||||||||||||||
Db   1648 CACAACTACCCCTGCCCCACGGGCCCGGGTGGTGTCTACACCGGCACCCCACGACAACGCACACC 1707

Qy   1198 ACCCTGGATGGTTCCGCACCGCGAGGCCGAGGCGGGCCTTCATGCGGGCTACCTG 1257
          ||||||||||||||||||||
Db   1708 ACCCTGGGCTGGTACCGCACGGCCACCCGCCACCCCCGAGGCCTTCATGGCGCGGTACCTG 1767

Qy   1258 GCCCGCTATGGCATCCCGTTGTTTGTCGGAATACGAGGTCGCGGCGCTTTGATCGAGCTG 1317
          ||| ||||||||||| | ||||
Db   1768 GCGGACTGGGGGGATCACCTTCCGGAAGAGAGGAGTGCCCTGGGGCCCTGATGCACCTG 1827

Qy   1318 GCCTTCAAAAGCCCGGCCAAGCTGGCTATTGTGCCTTTGAGGACGTGCTCGGGGCTGGGC 1377
          |||||||| |||||| |||
Db   1828 GGGATGAAGTCCGTGGCCCGGCTCGCCGGTCTACCCCGTGCAGGACGTGCCCTGGGGGGC 1887

Qy   1378 CCCGGAGCCCGCCATGAACTTCCCGCCGACGGCTGGGGGACAACTGGGCGTGGCGTACGCC 1437
          ||||||||||| |||||
Db   1888 AGCGAGGCCCGCGATGAACTACCCGGGAAGGCCCTCGGGGAACTGGGCGGCCTGCCGGCTCCTC 1947

Qy   1438 GAAGGCGACCTCGAGCCCGGTCTGCGCCGGGACTGCGGCCCTGGCCGGAGGCCAGCCAG 1497
          |||||||||||||||||||||||
Db   1948 CCGGGGGGAGCTTTCCCCGGAGCCGAGGCCGGTTAGGGCCATGGCCCGAGGGCCACGGAA 2007

Qy   1498 CG 1499
          ||
Db   2008 CG 2009

Qy 301 TGGCTTTACCAGACCCGCTCGTTGCGCGGGCTTTCGGGCAAGG 360
        ||||| |||||||||||||||||||| ||||| |||||
Db 295 CTCCTCTACGCCTGGAAGTGGCCGCCGTCTTGAGGCGCCCTTGCGCTGGGCGGTCCCTGCGCGGCTTCAAGGAAAAG 354

Qy 361 GCTTCGGCCCAGGATAAGACCCGACTGGAAGCCTTTATCGAGGCGCTTCTGGCTG 420
        |||| |||||||||||||||||||
Db 355 GCCTCCCCGGAGGAGGCGGAGGGCCTTGCCGCCTTCGCGGGAGGAGGGGAGGCCTGGTGTGCTC 414

Qy 421 GAAGACTATGCGGCTCTTTATGCCCTCAAGACCCGGTTTGACGGCAAGCCCTTGGAACGAG 480
        ||||||| |||||||||||||||
Db 415 GAGGACTACGCCCCTCTTCTTCATGGCCGTCGAGGGGCGCACGCGGGGGGCTTCCCTGGAACCGG 474

Qy 481 TGGAGCCCCGAGCTGCGGCACCGTGCGCCGGCCGGTGCCCTGGCCCAGGGCCCGTGAGGAGCTG 540
        ||| ||||||||||||
Db 475 TGGCCCCTTCCCGGCTGCCCTGGCCTTAGGGAGGCGGAGAGGAAAAAGCGCCTTG 534

Qy 541 GCCGGAGGGTGCGCCCTTACGAGTCGGATTCAGTGGCCTTTTTTATCTGGAATGGGGCCAG 600
        ||| ||||||||||||||||
Db 535 GCCGGAGGAGGTGGCCTTTCCACCGCCTGGCTCTTTCCCTTCCTCCGCGGGGGGGCC 594

Qy 601 ACCAAGGCCTATGCCGAATCCAAGGGGATTCAGATTATTATCGGCGCGATATGCCCATCTTTGTG 660
        ||| ||||||||||||
Db 595 TTGAAGGCGGAGGCGAGCGGTTGGGCATCCGGATCATCGGGGACATGCCCATCTTCGTG 654

Qy 661 GCCTTCGAATTCCTCAGATGTCTGGCCAACCCGCAGTACTTCTCACCTCGAGGCCGATGGC 720
        ||| ||||||||||||
Db 655 GCCGAGACTCCGCCGAGGTCTGGGCGCGTGGGCCCACCCCCAGTGGTTTCACCTGACGAGGAGGGGC 714

Qy 721 AACCCCACGGTGTGGCGGCGGTTCCCGCGGGACTACTTCTCCCGAAACCGGCCAGCTCTGG 780
        ||| ||||||||||||
Db 715 CGCCCCACGGTGTGGCGGCGGGGTGCCCCCGGGACTACTTCTCTGGAGAGGCGCTGGTCAAA 774

Qy 781 GGCAATCCGCTCTATCGCTGGGATGTGATGGAAAGGACAACTTTGCCTGGGTGCATTGCC 840
        ||| ||||||||||||
Db 775 GGCAACCCCTTTTACCGCTGGGGACGTTTGGGACGCGGAGGGGTTCTCTCTTCTGGATCCGC 834

Qy 841 CGGCATAAGGCAGTCGCTCAAGACAGTCGCCCTTGGAGCTTCTTCCACCTGGTGCGGCATGCGCGCGTGGGGTTT 900
        ||| ||||||||||||
Db 835 CGTCTGGAGAAGGGCCCTTGGTGCCCATTGTGTGGAGGGGCGCTCGGTCAAACCTTCCGGCGGCTTT 894

Qy 901 GAAGCCTACTGGGAGGTTCCGTTTGGCGGGATCCCCAATGCTGCCCCACGCGGGGACCTGCCCAAGCTGCCCCAG 960
        ||| ||||||||||||
Db 895 GAGGCCTACTGGGAGGAATCTGGGAGATGCGGTTGGGACGGGCGGTGGGCGTGGGGTCAAG 954

Qy 961 GCCCCCAGGGGAGAAGCTGTTTGCTGCGGTGCGCGGCGATGCGCCCAACTGAGCGATTCGCCCCATCATT 1020
        ||| ||||||||||||
Db 955 GCCCCGGGGGGAGAGAGCTCTTCCAGAGATCCAGGAGGTTCTTCGCGCGGAGGTCTTCGCGGTTC 1014

Qy 1021 GCCGAAGACCTGGGGTGATCACCCCCGAGGTGGAGGGCTTGCGGAGGCTTCGGGGTC 1080
        ||| ||||||||||||
Db 1015 GCCGAGACCTGGGGTCATCACCCCGAGGTGGAGGCGCCTGCGCCGACCGCTTCGGCCTT 1074

Qy 1081 CCCGGTGAGGACATGAAGAGATTTGCAGTTTGCTTTT---TCCGGTGAGGACGACAACGCCTTTTTGCCC 1137
        ||| ||||||||||||
Db 1075 CCCGGGGATGAAGGTGAAGTTCCTGCAGTTCGCCCTTTGACGACGGGATGGAAAAACCCCCTTCCTCCCC 1134

Qy 1138 CACAACTACCCCGCGCACGGCAATGTGGTGTATACAGCGGGACCCAGCAACGACACC 1197
        ||| ||||||||||||
Db 1135 CACAACTACCCCTGCCACGGTCCGCCGGGGTGGTCTTACACCGGCACCGACAACGACACC 1194

Qy 1198 ACCCTGGGAGTGGTTCCGGACCGGAGCCCGGAGGGCCGAGCGGCCGGCCTTCATGGGCGGGCTACCTG 1257
        ||| ||||||||||||
Db 1195 ACCCTGGGGCTGGTACCGCACGGCCACCCCCCCAGAGAAGGCCTTCATGCGCGCGTACCTG 1254

Qy 1258 GCCGCGTATGGCAATCCGTTGTGTTGTTTCGGAATACAGACGTCGCGGGCGCTTTGATCGAGCTG 1317
        ||| ||||||||||||
Db 1255 GCGGACTGGGGGATCACCTTCGGGAAGAGAGGAGGTGTCCCCTCGGCCCCTGATGCACCTTG 1314

Qy 1318 GCCTTCAAAAAGCCCGGCCAAGCTGGCTATTGTGCCTTTGCCTGTGCTGTCTGGGCTGGGC 1377
        ||| ||||||||||||
Db 1315 GGGATGAAGTCCGTGGCCGCGCCGGCTGCCCGTCTACCCGGTCGCAGGACGCTGCCGGCGCTCGGGC 1374

Qy 1378 CCCGGAGGCCCGGACCATGAACTTCCCCGGACGAACTGGCGCTGGGGGCAACTGGCGGCTACGCC 1437

Db 1375 AGCGAGGCCCGGATGAACTAACTACCCGGAAGGCCCTCGGGGAACTGGGCCTGGCGGCCCAGCCAG 1434

Qy 1438 GAAGGCGACCTGAGCCGCGTCTGCGGGACTGGCGCCGGGGGCGACTGCCGGACTGCCGGCGCAGCCGGCCAGGCCAG 1497
        ||| ||||||||||||
Db 1435 CCGGGGGAGCTTTCCCCGGAGCACGGGGCGAGCGCTTAGGGCCATGGCCGAGGCCACGGAA 1494

Qy 1498 CG 1499
        ||
Db 1495 CG 1496

RESULT 5
ADU07728
ID   ADU07728 standard; DNA; 1539 BP.
XX
AC   ADU07728;
XX
DT   13-JAN-2005 (first entry)
XX
DE   DNA sequence #221 encoding amylase.
XX
KW   Amylase; glucoamylase; 1,4-alpha-D-glucan glycohydrolase; alpha-amylase;
KW   exoamylase; beta-amylase; glucosidic bond hydrolysis; starch; sugar;
KW   glucose; maltodextrin; thermostable; gene; ds.
XX
OS   Unidentified.
XX
PN   WO2004091544-A2.
XX
PD   28-OCT-2004.
XX
PF   08-MAR-2004; 2004WO-US007096.
XX
PR   06-MAR-2003; 2003US-00385305.
PR   28-MAR-2003; 2003US-0459014P.
XX
PA   (DIVE-) DIVERSA CORP.
XX
PI   Callen W,   Richardson T,   Frey G,   Gray K,   Kerovuo JS,   Slupska M;
PI   Barton N,   O'donoghue E,   Miller C;
XX
DR   WPI; 2004-775532/76.
DR   P-PSDB; ADU07729.
XX
PT   Novel recombinant amylases, useful for hydrolysis of starch into sugars,
PT   in detergent composition, for producing high-maltose or high-glucose
PT   syrup, in brewing or alcohol production, bio-bleaching composition of
PT   paper or pulp product.
XX
PS   Claim 4; SEQ ID NO 438; 436pp; English.
XX
CC   The invention relates to the isolation of polypeptides with amylase
CC   activity, and polynucleotide sequences encoding the polypeptides. The
CC   amylase activity comprises a glucoamylase activity, 1,4-alpha-D-glucan
CC   glycohydrolase activity, alpha-amylase activity, exoamylase activity or
CC   beta-amylase activity. The amylase activity comprises hydrolyzing
CC   glucosidic bonds in a starch to produce sugars such as glucose and
CC   maltodextrins. The amylase activity comprises cleaving a maltose or a D-
CC   glucose unit from the non-reducing end of the starch. The amylase
CC   activity is thermostable and the polypeptides of the invention are active
CC   over a wide range of temperatures. The polypeptide sequences are useful
CC   for making anti-amylase antibodies, for identifying a modulator of an
CC   amylase activity, or identifying an amylase substrate. The polynucleotide
CC   sequences are useful for producing recombinant polypeptides, and for
CC   modifying codons in a nucleic acid encoding a polypeptide with an amylase
CC   activity to increase or decrease its expression in a host cell. The
CC   polypeptide and polynucleotide sequences are useful for making or
CC   modifying a small molecule, and for determining a functional fragment of
CC   an amylase enzyme. Primers to the polynucleotides are useful for
CC   amplifying them. This sequence represents a polynucleotide sequence of
CC   the invention.
XX

SQ   Sequence 1539 BP; 385 A; 401 C; 378 G; 375 T; 0 U; 0 Other;

Query Match   25.1%; Score 378; DB 13; Length 1539;
Best Local Similarity   55.7%; Pred. No. 3.3e-80;
Matches 801; Conservative   0; Mismatches 610; Indels   27; Gaps   3;

```
Qy      1 ATGCAACTCCAACGCGGCTTTTGGAATTTGCTCCACCCACCAGTTTCCGGGTCGCTGG 57
            ||||| ||  || |||||||||   ||||||||||||||||||||
Db      1 ATGAAATTTAAGCGGCGGCGGGGAATCATCCTTCATCCTACCACCGGTTCGCGGGCCGGAT 60

Qy     61 GGGATTGGGGCTCTGGGCCGCGAGGCCGAGCCGTTTTGGACTGGCTGGCCGGATGCGGGGA 120
            |||| | |||  |||||||||||||||||||
Db     61 GGAATTGGTGACCTTGGACCGGCATACCCGCTGGATCAACTTTCTAGCCGGAATCCGGC 120

Qy    121 GCCCGCGCTGGTGGTGGCAGGTCTTACGGTCGGCCCTACCAGTTACGGCGACTCGCCGTACCAG 180
            |||||||| || |||||||||| ||||||||||||||||
Db    121 TGCAAACTATGGCAGGTGCTGCCGCGTTGCGCGACTCGCCCTATCAG 180

Qy    181 TCCTTCTCCGGCTTTTGGCCGGTAACCCGTATTTGGTTGACCCCGAGATGCTGATTGAAAAA 240
            ||||| | |||||||||| |||  |||||||||  |||
Db    181 TGTTTCTCAGCATTGCCGGTAATCCCTATCTGGTCAGCCCGCGCGTGTTGTTGCTGGAAAAC 240

Qy    241 GGCTGGCTGGAACAAAGCGA------AGCGCCCCGCGTATCCGACCCCAGCGCGTGGAT 294
            ||| |||||||| ||        ||| |||||||| |||
Db    241 GGTTTGGTTACGGCGGGGAGTGATCTGACCGACCGGCCTGAATTTAATCCAATTCGAATTGAT 300

Qy    295 TATGGCTGGCTTTTACCAGACCCGCTGGCCCGTGTTGCGGCGGCTTTCGGCGGGGGTTTCGG 354
            ||||| |||||||| ||
Db    301 TACGGTTGAAGCCATCACTTGAAATTAAAAATACTGATCGCGCTTACGAACAAACTTTAAG 360

Qy    355 GCAAGGGGCTTCCGGCCAGGATAAGACCCGACTGGAAGCCTTTATCGAGGCGCTTC 414
            ||  ||
Db    361 CGCCTCAATAATCCTGAACTAAAAATATGAATATCAGCAATTCACGGAAAATGAGACGGAC 420

Qy    415 TGGCTGGAAGACTATGCGCTCTTTATGGCCCCTCAAGACCCCGGTTTGACGGCAAGCCCTGG 474
            ||| || |||||||| ||| |||  |||||||
Db    421 TGGTTGAATGATTTTGCCCTTTTATGGCCATTAAGAGGCTAACGGGGGGGTATCCTGG 480

Qy    475 AACGAGTGGAGCCCCGAAGTTGCGCGGACCGGCTGCGCCCTGAGACCGGCCCGTGAACCCGTGAG 534
            || ||||||| |||||||||||| |||||||
Db    481 GACAACTGGCCAAAGGAACTGCGTAGCCGACCGGCCAATCGAAGTTCAAACAA 540

Qy    535 GAGCTGGCGCGGAGGAGGTGGCCCTTTACGAGTGGATTCAGTGGCTTTTATCTGGAATGG 594
            ||||||||  |||||||| |||  |||||
Db    541 ACCGAAGCCGACCGCCATTCAACGCCATCAACGCCTTTGCACAGTTTTTATTCTTCCGCGCGTGG 600

Qy    595 GGCCAGACCAAGGGCTATGCGCGAATCCAAGGGATTCAGATTATCGGCGATATGCCCATC 654
            |||| ||| || |||||||||
Db    601 CTTGACCTGAAAGCCTCACGCCAGGCGCGTTGTGCACCAATCAAAATCATCGGTGATATACCGATT 660

Qy    655 TTTGTGCCGGCCTTCGATTCCTCCGATGTGTCTGGCGTCTCGGCGCAAGTCACTACTTCTCAGGGCC 714
            |||||||| |||||  |||||||||||||| | ||||||  ||||
Db    661 TTTGTGCTTACGATAGCGCAGATAGCGTTTGGTCAGACCCCAGAGTTGTTTATCTGAATGAA 720

Qy    715 GATGGCAACCCACCGGTGGTGCGGCGGGTTCGGCGGGGGCTTCGGCGATATGCCAG 774
            || |||||| |  ||| ||| ||| |||||||||||  ||||||
Db    721 GAAGGCAAACCAACTGTTGTGCAGCGCGTTCCACCACGGGTCAA 780

Qy    775 CTCTCGGGCGCAATCCCGCTCTATCGCTGGGATGTGATGGAAAGGGACAACTTTGCCTGGTGC 834
            |||| |||  |  |||||||||||| ||||||| |||
Db    781 TTATGGGTAACCCCGTTTACAAGTGGGAAGTTCATCGTCAGCAGATTTCCGGGATATTTCGC 840

Qy    835 ATTGCCGCCGCATAAGGACAGTCGCTCAAGCAGTCGTGCGCATCGACCACTTCCGC 894
            |||||||||||  |  || ||  |||||| ||||| ||||
Db    841 ATTCCGCCGGATGCAAGCCACATTACGGCTGGTAGATATTGTCCGCCCTGACCATTTCGC 900

Qy    895 GGGTTTGAAGCTACTGGGAGGTTCGTTTGCCCGCGCCAATGCTGTGGAGGGGCCTGG 954
            || |||| || |||||||| |||||||  ||||||
Db    901 GGATTTGCCGGTTATTGGGAGGAGTGAATCGCCCACCCTTTGGGATCGCCGCGCAAATTGGGCCGCTGG 960

Qy    955 GTCAAAGCCCCAGGGGAGAAGCGTGTTTGCTGCGGTGCGGGCCCAACTGAGGCGATGCGCCC 1014
            ||||||||||| ||| ||  || ||||  |
Db    961 GTAAAAGGGGCCGCCGGCCAAAGAGAATTATTCAATGCCATTCGGGATGCCCTTGGCGATTTTACCC 1020

Qy   1015 ATCATTGCCGAAGACCTGGGGGTGATCACCCCCGAGGTGGAGGCTTTCGGCGATGGCTTC 1074
            |||||||||||||| ||||||||||||| ||||||||||| ||||||||||||
Db   1021 ATCATCGGCTGAAGATTTGGGGGTCATCACCACCCCCATGTGATTTCCGGTGAGGAACAACGGCCTTTTG 1080

Qy   1075 GGGTTCCCGCGTGAAGAGATTTTGCAGTTGCTTTTCCGGTGAGGAGGAACAACGGCCTTTTTTG 1134
            || ||  |||||  ||||||||| |||||| || || ||||||| |||
Db   1081 AACTTGCCCGGTATGAAGGTGTTCCAGTTGCTTTTACAACGGACCCGCTCGATCCCTTC 1140

Qy   1135 CCCCACAACTACCCCGCACGGCAATGTGGTGGCTACAGCGGAGCCACGACCAACGAC 1194
            ||  || | ||||| |
Db   1141 CTGCCGCCACAACTACG---AAACCAATTGTGTAGCCTACCACCGACCAACGAC 1197

Qy   1195 ACACCCTGGGATGGTTCCGACCACGGCCGGCGAGCGGCCCTTCATGCGGCGCCTAC 1254
            ||||||||||| || |||| ||||||| ||||||||| |||||
Db   1198 ACCGCCCTGGGTGGTATCAAACGGCCACCGGAAAAAGACGTGACTTTATTCGGCGATAT 1257

Qy   1255 CTGGCCGCGTGCTATGGCATCCGTGTTGTTGTCGGAATACGAGGTCGGCGGGCTTTGATGCGAG 1314
            ||||||| ||   |||||||  || |||| ||||||       ||| ||||||||||| ||
Db   1258 CTGCGCCGTTCCGGC------------------GAAGATATATCATGGACGATGATCCGC 1299

Qy   1315 CTGGCCTTCAAAAGCCCGGCCAAGCTGCTATTGTGCTTTGAGGACGTGCTCGGGGCTG 1374
            || ||  |||| |||||| || ||||||||| || ||||||||
Db   1300 GCTGTGTGGTCTTCAGTGCCGTTTTGGCCGCCATGCAAGATATTTGAGCCCTT 1359

Qy   1375 GGCCCGAGGCCCGCATGAACTTCCCCGGACACGGCTGGGGGACAACTGGGCGTGCGCGCT 1432
            ||||||| ||||||| |||| |||||  |||||||||| ||||| ||
Db   1360 GGAAATGAAGCCCGCATGAAATCCACCGGGAAGAGCCACCGCGGCCAACTGGTTGTTGGAGAT 1417
```

RESULT 6
ADQ15025
ID   ADQ15025 standard; DNA; 1518 BP.
XX
AC   ADQ15025;
XX
DT   23-SEP-2004 (first entry)
XX
DE   DNA encoding Synechocystis sp. PCC 6803 amylo maltase polypeptide.
XX
KW   Amylo maltase; transgenic; foodstuff; food additive; modifier;
KW   saccharide; (alpha)-1,4-glucan; cooked rice; Japanese confectionery;
KW   snack; noodle; Chinese dumpling; shao-mai; fishery kneading;
KW   refrigerated; baby food; pet; animal feed; drink; food supplement;
KW   cyclic glucan; Synechocystis sp. PCC 6803; gene; ds.
XX
OS   Synechocystis sp. PCC 6803.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..1518
FT   /*tag= a
FT   /product= "Synechocystis sp. PCC 6803 amylo maltase
FT   polypeptide"
XX
PN   JP2004187674-A.
XX
PD   08-JUL-2004.
XX
PF   12-SEP-2003; 2003JP-00322319.
XX
PR   26-NOV-2002; 2002JP-00342966.
XX
PA   (EZAK) EZAKI GLICO CO LTD.
PA   (NIDE) NEC CORP.
XX
DR   WPI; 2004-528719/51.
DR   P-PSDB; ADQ15026.
XX
PT   New mutated amylo maltase polypeptide having increased enzyme activity
PT   and reduced hydrolyzing activity, useful for manufacturing foodstuffs
PT   such as Japanese confectionery, noodles, baby foods, and food additives.
XX
PS   Example 9; SEQ ID NO 16; 51pp; Japanese.

XX
CC   The invention relates to a novel amylo maltase polypeptide comprising the
CC   amino acid sequence of a wild-type amylo maltase with a substitution,
CC   addition or deletion at a position, where the amino acid residue
CC   interacts with acarbose or amino acids other than the wild-type amylo
CC   maltase amino acid sequence. The invention further comprises: a nucleic
CC   acid molecule containing a nucleic acid sequence encoding the novel amylo
CC   maltase polypeptide; a vector containing the nucleic acid sequence; a
CC   cell containing the nucleic acid sequence; biological tissue containing
CC   the nucleic acid sequence; a transgenic organism containing the nucleic
CC   acid sequence; a foodstuff, a food additive or a modifier of foodstuff
CC   containing the nucleic acid sequence; a computer readable recording
CC   medium which contains the information of the nucleic acid sequence
CC   encoding the amino acid sequence of the novel amylo maltase polypeptide;
CC   and a cyclic glucan obtained by reacting the novel amylo maltase
CC   polypeptide on saccharides which have a linear structure of (alpha)-1,4-
CC   glucan. The novel amylo maltase polypeptide is useful for manufacturing
CC   foodstuffs such as cooked rice, Japanese confectionery, snacks, bakeries,
CC   noodles, Chinese dumpling, shao-mai, fishery kneading goods,
CC   refrigerating foods, baby foods, pet foods, feed for animals, drinks, and
CC   food supplements, where the method involves adding the novel amylo
CC   maltase polypeptide to the foodstuff before or immediately after heat
CC   processing of the raw material, where the novel amylo maltase polypeptide
CC   generates a cyclic glucan from the starch of the foodstuff. The novel
CC   amylo maltase polypeptide is useful for manufacturing glucan which has a
CC   cyclic structure of (alpha)-1,4-glycoside linkage, foodstuff materials
CC   and food additives, modifier of foodstuffs, a drink or eatable
CC   composition, infusion solution or composition for adhesion, where the
CC   method involves reacting the saccharide containing the linear structure
CC   of (alpha)-1,4-glycoside linkage or its derivatives with the novel amylo
CC   maltase polypeptide. This polynucleotide sequence represents the DNA
CC   encoding a Synechocystis sp. PCC 6803 amylo maltase polypeptide of the
CC   invention.
XX
SQ   Sequence 1518 BP; 362 A; 392 C; 390 G; 374 T; 0 U; 0 Other;

Query Match   21.9%; Score 329.2; DB 12; Length 1518;
Best Local Similarity   53.3%;
Matches 800; Conservative   0; Mismatches 673; Indels   27; Gaps 4;
                                   Pred. No. 1.5e-68;

[DNA sequence alignment blocks, Qy/Db pairs, positions 1–534]

RESULT 7
ADK45052
ID   ADK45052 standard; DNA; 1518 BP.
XX
AC   ADK45052;

[DNA sequence alignment blocks, Db/Qy pairs, positions 481–1485]

XX
DT   24-FEB-2005 (first entry)
XX
DE   Streptococcus pneumoniae gene, Seq ID No 1567.
XX
KW   ds; gene; Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.
XX
OS   Streptococcus pneumoniae.
XX
PN   US6699703-B1.
XX
PD   02-MAR-2004.
XX
PF   26-MAY-2000; 2000US-00583110.
XX
PR   02-JUL-1997;   97US-0051553P.
PR   12-MAY-1998;   98US-0085131P.
PR   30-JUN-1998;   98US-00107433.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CE;
XX
DR   WPI; 2004-212399/20.
DR   P-PSDB; ADK47713.
XX
PT   New nucleic acid molecules and polypeptides useful for diagnosing,
PT   preventing and treating pathological conditions resulting from bacterial
PT   infection, e.g. Streptococcus pneumoniae infection, and in drug
PT   screening.
XX
PS   Disclosure; SEQ ID NO 1567; 301pp; English.
XX
CC   The invention relates to isolated Streptococcus pneumoniae nucleic acids
CC   and polypeptides. The nucleic acids and proteins are useful for
CC   diagnosing, preventing and treating pathological conditions resulting
CC   from bacterial infection, such as S. pneumoniae infection. These may also
CC   be used for drug screening procedures. The present sequence represents a
CC   Streptococcus pneumoniae nucleic acid of the invention. Note: The
CC   sequence data for this patent did not appear in the printed specification
CC   but was obtained in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 1518 BP; 435 A; 308 C; 351 G; 424 T; 0 U; 0 Other;

Query Match   19.2%; Score 289; DB 13; Length 1518;
Best Local Similarity   51.7%; Pred. No. 6.2e-59;
Matches 764; Conservative 0; Mismatches 695; Indels 18; Gaps 4;

Qy   11 AACGCGCTTTTGGAATTTGCTCCACCCCACCAGTTTCCGGGTCGCTGGGGGATTGGGG 70
          |||| || |||| ||| ||||| ||||| ||||| |||||||| ||
Db   8 AACGTCAAAGTGGTGTTGATGCATCTCTTCTCTTCCAGGAGTTTACCGAAATCGGAT 67

Qy   71 CTCTGGGCCGGCCGGAGGCCCGAGCCGTTTTGGACTGGCTGGCCGGATGCCGAGCCGCTGGT 130
          |||| | ||| |||| | |||||||
Db   68 CATTTGGTCAAAGTGCTTACGACTTCCGTTCGGTGTCTGGGGATTCTCGGTTACCAATCTTTACC 127

Qy   131 GGCAGGTCTTACCGCTGGGCCCTACCAGTTACGGCCGACTCCGCGTACCAGTTCCTTCGG 190
          |||||| |||| |||||| ||
Db   128 GGCAAATCCTTCCATTAGGAGCAACTAGTTACGGGGATTCTCCTTACCAATCTTTCTCAG 187

Qy   191 CTTTTGCCGGTAACCCGTATTGGTTGACCCGGAGATGCTGATTGAAAAAAGGCTGGCTGG 250
          || | ||||||| ||| ||||| || |
Db   188 CCTTCGCAGGAAACACTCATTTATCGATTTAGATATATCTTGGTGGAGCAAGGTTTGTTGG 247

Qy   251 AACAAAGCGAAGGCCCCCCGCTATCCGACCCAGCGCGTGAATATTGGATTATTGACGTTACC 310
          ||||| | || |
Db   248 AGCAAGTGACCCTGAAGGAGTTGCGGTAGCCAATGCGTCTCAAAGTTGACTATGCTA 307

Qy   311 AGACCCGCTGGCCCCTGTTGCCGCGGGTTTCGGGCAAGGGCTTCGGCCC 370
          ||| | | |||
Db   308 AAATCTACTACGCACGTCGTCCTCTTTAGAAAAAGCGGTGAAACGTTTCTTTGAAGTCG 367

371 AGGATAAGAACCCGACTGGAGAGCCTTTATCGAGGCCGAGCGCTTCTGGCTGGAAGACTATG 430   Qy
    ||| ||||||
368 GAGATGTTAAAGATTTTGAGAAATTGCTCAAGACAACCAATCATGGCTTGAGCTCTTTG 427   Db

431 CGCTCTTTATGGCCCTCAAGACCCGGTTGACGGCAAGCCTGAACAGTGG---AGCC 487   Qy
    |||||
428 CTGAGTATATGGCTATCAAAGAGCATTTTGACAATCTTGCTTGGACTGGAATGGCCAGATG 487   Db

488 CCGAGCTGCGCGACCGGCTGAACCGGCCCGTGAGGAGTCCCGTGAGCGCGAGG 547   Qy
    ||||| |||||||
488 CAGATGCTCGTGCTCGTAAAGCTTCAGCACTTGAAAGCTATCGTGACGCAGACA 547   Db

548 AGGTGGGCCCTTTACGAGTGGGATTCAGTCAGTGCTTTTTATCTGGAATGGGCCAGACCAAGG 607   Qy
    ||||| ||||||
548 AGTTGGTTTTACCACCGGTGTCCGTGGGGATTCAGATTATCGGCGGATTATCGGCTCGGGCAATC 607   Db

608 CCTATGCCGAATCCAAGGGATTCAGATTATCGGCGGATTATGCCCATCTTGTCGCCTTCG 667   Qy
    |||||||
608 CTTACGCTAACGACAACCACATCGAAATCGTTGGGGACATGCCAATCTACGTAGCGCGAAG 667   Db

668 ATTCCCTAGCAGATGTCTCGGGCCAACCCGCAGTACTTCTACCCTGCAGGCGATGCGCAACCCCA 727   Qy
    ||||| ||| ||||||
668 ATTCAAGTGATATGTGGGCAAATCCACATCTCTTCAAAACAGATGTCAATGGTAAGGCTA 727   Db

728 CGGTGGTGGCGGGCGTTCCGCGGGACTACTTCCTCCGGGAAACCGGCCAGCTCTGGGGCAATC 787   Qy
    |||||
728 CTTGTATCGCAGGATGCCACCAGATGAGTTTCTGTAACTGGTCGTAACTGGTTTGGGGTAATC 787   Db

788 CGCTCTATCGCTGGGATGTGATGGAAAGGGACAACTTTGCCTGGTGCATTGCCCGCATAA 847   Qy
    ||||| ||
788 CAATCTATGACTGGGAAGCAACAATGGACAAAGACGACGCTACAAATGTGAGATTGAACGCTTGC 847   Db

848 GGCAGTCGCCTCAAGCAGTGCCACCTGTGCGCATCGACGACCACTTCCGCGGGTTGAAGCCT 907   Qy
    ||||| || || || ||| ||||||
848 GTGAAAGCTTCAAAATCTACGATATCGTTCGTATGCGACCACTTCCGTGCGGCTTCGAATCTT 907   Db

908 ACTGGGAGGTTCCGTTGGCCGGCCCAATGCGTGGAGGGGCGGCTGGGTCAAAGCCCCAG 967   Qy
    ||||||| || |||||
908 ACTGGGAAATCCCTGCGTTCCGATACAGCAGCACCTGGTGAGTGAGTGAAAGGTCCAG 967   Db

968 GGGAGAAGCTGTTGCTGCCGGTGCGGCGGCCCAACTGAGCGATGCGCCCATCATTGCCGAAG 1027   Qy
    ||||| || ||| ||
968 GCTACAAGCTTTTGCAGCGCGCCGTTAAGGAAGAACTTGGTGAGCTAAACATCATCCGCAGAG 1027   Db

1028 ACCTGGGGTGATCACCCCCGAGGTGGAGGCTTTGCGCGATGGCTTCGGGTTCCCCCGGCA 1087   Qy
     ||| ||||||
1028 ACCTTGGCTTCATGCAGCATCTCATGCAAGTGATCGAATTGCGTGAACGTACTGGCTTCCCAGGAA 1087   Db

1088 TGAAGATTTGCAGTTTGCTTTTCCGGTGAGACAACGCCTTTTGCCCACAACTACC 1147   Qy
     ||||||||| ||
1088 TGAAGAATTCTTCAATTTGCCTTCAACCCAGAGACGAAAGCATTGATAGCCACACT--- 1144   Db

1148 CCGGCACGGCAAATGTGGTGTACGGTGTACAGCGGAACCCACGACACAGCCTGGGAT 1207   Qy
     ||||
1145 TGGCACCTGCTAACTCAGTTATGTACACAGGAACACACAAATAGGTTCTTGGTT 1204   Db

1208 GGTTCCGCACCCGCCCGAGGCGGGCCTTCATGCGGGCCGTAGCCGTATG 1267   Qy
     ||||||
1205 GGTACCGTAATGAGATTGATGATGCG----------ACTCGGTGAGTACATGGCTCGTTACA 1255   Db

1268 GCATCCGTGTTGTTGTCGGAATACGAGGTCGCGGGCGTCTTGATCGAGCTGGCCTTCAAAA 1327   Qy
                   |||
1256 CGAACCGTAAAGAATACGAAACAGTGGT---ACACGCTCTATGCTACAGTATTTCAT 1312   Db

1328 GCCCGGCCAAGCTGGCTATTGTGCCTTTGCAGGACGTGCTGGGGCTGGCCGAGGCCC 1387   Qy
     ||||||
1313 CAGTTAGCTTTATGGCAATTGCAACTATGCAAGATTTACTAGAATTGGATGAGGCAGCTC 1372   Db

1388 GCATGAACTTCCCCGGACGGCTGGGGGACAACTGGGGCGTGGGCGTGACCACCGGAGGCGACC 1447   Qy
     |||||| | |||
1373 GTATGAACTTCCCATCTACCCTTGGTGGTGAAACGGGTGAAACTGGTCTTGGCGTATGACTGAAGATCAAT 1432   Db

1448 TCGAGCCCGGTCGGGCGCGGGACTCGGGGCCCTGGC 1484   Qy

```
Db   1433 TGACACCAGCTCTGCGAGGAAGGTTTGCTTGACTTGAC 1469
          |||     ||||||   |||  |||  ||| |  ||| |
```

RESULT 8
ABX07913
ID   ABX07913 standard; DNA; 1515 BP.
XX
AC   ABX07913;
XX
DT   27-OCT-2003 (revised)
DT   11-FEB-2003 (first entry)
XX
DE   S. pneumoniae type 4 strain coding region #2201.
XX
KW   Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;
KW   ear infection; antiinflammatory; antibacterial; immunostimulant;
KW   auditory; respiratory; gene therapy; vaccine.
XX
OS   Streptococcus pneumoniae; type 4 strain.
XX
PN   WO200277021-A2.
XX
PD   03-OCT-2002.
XX
PF   27-MAR-2002; 2002WO-IB002163.
XX
PR   27-MAR-2001; 2001GB-00007658.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Masignani V, Tettelin H, Fraser C;
XX
DR   WPI; 2003-040579/03.
DR   P-PSDB; ABU02622.
XX
PT   New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT   useful as medicaments for treating or preventing a disease or infection
PT   due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
PT   ear infection.
XX
PS   Claim 6; SEQ ID NO 4401; 56pp; English.
XX
CC   The invention relates to a protein comprising or having at least 50%
CC   identity to any of the 2469 amino acid sequences, identified in the
CC   specification (available on a computer readable format), or its fragment,
CC   expressed from 2469 of 2489 identified DNA coding regions from the
CC   Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC   ABS56454. Also included are an antibody which binds one of the proteins,
CC   treating a patient by administering the protein, DNA or antibody (in a
CC   composition), a kit comprising first and second primers, which are the
CC   nucleic acid cited above or fragments between nucleotides 8-100 of a
CC   sequence not defined in the specification, for amplifying a target
CC   sequence contained within a Streptococcus nucleic acid sequence, where
CC   the first primer is substantially complementary to the target sequence
CC   and the second primer is substantially complementary to the complement of
CC   the target sequence, and where the parts of the primers having
CC   substantial complementarity define the termini of the target sequence to
CC   be amplified, assay comprising contacting a test compound with the
CC   protein, and determining whether the test compound binds to the protein
CC   and a Streptococcus pneumoniae bacterium, where one or more genes
CC   encoding the proteins has been rendered inactive. The proteins, nucleic
CC   acid molecules, antibody and compositions are useful as medicaments for
CC   treating or preventing a disease or infection due to streptococcus
CC   bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
CC   media or ear infection. They are also useful in developing vaccines,
CC   diagnostics and antibiotics. The methods are useful for identifying
CC   immunodominant proteins. The present sequence is one of the 2489
CC   identified coding region from the genomic sequence. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences. (Updated on 27-OCT-2003 to

CC   standardise OS field)
XX
SQ   Sequence 1515 BP; 433 A; 306 C; 351 G; 425 T; 0 U; 0 Other;

Query Match   19.1%;   Score 287.4;   DB 10;   Length 1515;
Best Local Similarity 51.7%;   Pred. No. 1.5e-58;
Matches 763; Conservative 0; Mismatches 696; Indels 18; Gaps 4;

```
Qy    11 AACGGCCTTTTGGAATTTGCTCCACCCCACCAGTTTTCCGGTCGCTGGCGGGGATTGGGG 70
            || |||  |||| |||||  ||   ||    ||||     |  | || | |
Db     8 AACGTCAAAGTGGTGTGTTGATGCACATCTCTTCCTTCCAGGAGCTTACGGAGCGAATCGGAT 67

Qy    71 CTCTGGGCGCCGCAGGCCGAGCCGTTTTGGACTGGCTGGCCGATGCGGAGCCCGCTGGT 130
            | || |  |  |  ||| |    |||||| || |   ||| |
Db    68 CATTTGTCAAAGTGCTTACGACTTCGTTGATTTCTTGGTCGTACAAAACAACGTTACT 127

Qy   131 GGCAGGTCTTACCGCTGGGCCTACCAGTTACGGCGACTCGCCGTACCAGTCCTTCTCGG 190
            |||    ||| ||   ||||  || |||||| ||| |   |||   |  | || | |
Db   128 GGGCAAATCCTTCCATTAGGAGCAACTAGTTACGGGGGATTCTCCTTACCAATCTTTCTCAG 187

Qy   191 CTTTTGCCGGTAACCCGTATTTGGTTGACCCCGAGATGCTGATTGAAAAAGGCTGGCTGG 250
            | || || |  |||  |  ||| ||| |     ||| |   |  | |    | |
Db   188 CCTTCGCAGGAACACTCATTTTATCGATTTAGATATCTTGGTGGAGCAAGGTTTGTTGTGG 247

Qy   251 AACAAAGCGAAGCGCCCCGCGTATCCGACCCAGCGGCGTGATTATGGCTGGCTTACC 310
            || ||    || | |  ||||  | ||| | ||| | || |  | |||   ||| | |
Db   248 AAGCAAGTGACCTTGAGAGGAGTTGGTAGCGAGTGCGTCTGAAGTTGACTATGCTA 307

Qy   311 AGACCGCTGGCCCCTGTTGCCGCGGGCTTTCGCGCGGGGTTTCGGGCAAGGGCTTCGGCCC 370
            | || |  |        |  || |  |  ||    | || ||   | ||| |  | |   |
Db   308 AAATCTACTATGCACGTGCGTCCTCTTTAGAAAAAGCCGTGAAACGTTTCTTTGAAGTCG 367

Qy   371 AGGATAAGACCCGACTGGAAGCCTTTATCGAGGCGGAGCGCTTCTGGCGTGGAAGACTATG 430
            ||||   | ||| | ||| |    |||  | ||| |||    |    |||    ||| |||
Db   368 GAGATGTTAAAGATTTTGAGAAATTTGCTCAAGACAACAACCAATCATGGCTTGAGCTCTTTG 427

Qy   431 CGCTCTTTTATGGCCCTCAAGACCCGGTTGACGGCAAGCCCTGGAACGGTCG---AGCC 487
            | || ||| ||||| |||    ||   ||| | |   |   |||||   |      |||
Db   428 CTGAGTATATGGCTATCAAAGAGTATTTTGACAATCTTGCTTGGACTGAATGGCCAGATG 487

Qy   488 CCGAGCTGCGCGACCGTGAACCGGCTGCCCAGGCCCGTGAGGAGCTGGCCGAGG 547
            | || ||    ||| ||||||| |||| |||| || |||| || ||| | |
Db   488 CTTACGCTAACGACAACCACATCAACATCTCTCAAAACAGATGTCAATGGTAAGGCTA 547

Qy   548 AGGTGGCCCTTTACGAGTGGGATTCAGTGGCTTTTATCTGGAATGGGCCCAGACCAAGG 607
            ||  ||   |||| | || |  || |  || |||| ||  |||  ||    ||| | |
Db   548 AGTTGGTTTACCACCGTGACTCAATACTTCTTCTTCTTCCAACAATGGTTGAAATTGAAAG 607

Qy   608 CCTATGCCAATCCAAGGGGATTCAGATATCGGCGATAATGCCCATCTTGTCGCCTTCG 667
            | || || || | ||  ||| || | || | | ||||| || || || | | |  | |
Db   608 CTTACGCTAACGACAACCACATCATCGAAATCGTTGGGACATGCCAATCTACGTAGCGGAG 667

Qy   668 AATTCCTCAGATGTCTGGCCAACCCAGTACTTCTACCTCGAGCGCCGATGCGAACCCCA 727
            | ||    ||| | |||| || || |  | | || ||    ||    |||    ||  |
Db   668 ATTCAAGTAGATATGTGGGCAAATCCACATCTCTTCAAAACAGATGTCAATGGTAAGGCTA 727

Qy   728 CGGTGGTGGCGGCGTTCCGGCGGGACTACTTCTCCGGGAACACCGGCTGCTCTGGGCAATC 787
            |  ||  |  |  | || ||| | ||  |||| |    ||  | | |     ||||  |||
Db   728 CTTGTATCGCAGGATGCCCACCAGATGAGTTTTCTGTAACTGTCAGCTTCAGCTTTGGGGTAATC 787

Qy   788 CGCTCTATCGCTGGGATGTGATGGAAAGGGACAACTTTGCCTGCGTGCATTGCCCGCATAA 847
            |  ||| | | |||| | ||| |||| |   || | | |    ||    | |   |  ||
Db   788 CAATCTATGACTGGGAAGCAATGGAAGCGTACAAAATGTGATTGAACGCTTGC 847

Qy   848 GGCAGTGCGCTCAAGCAGTGCCACCTGGTGCGCATCGACCACTTCGCGCGGTTTGAAGCCT 907
            |  |||   |  |||  |||  ||| || |||  || | ||||||||| |  || |||| ||
Db   848 GTGAAAGTTCAAAATCTCGTGGTTCCGATACAGCAGCACCACTTCCGTGCCTTCGAATCTT 907

Qy   908 ACTGGGAGGTTCCGTTCGGCCGGGATGCCCAATGCTCGTGGAGGGGCGGCTGGTCAAAGCCCCAG 967
            ||||||   | ||| ||||||  ||| | ||| || ||| | |||  ||  |  | || |  |||
Db   908 ACTGGGGAAATCCCTGCTGGTTCCGATACAGCAGCACCTGGTGAGTGGGTGAAAGGTCCAG 967

Qy   968 GGGAGAAGCTGTTTGCTGCGGTGGCGGGGCCCATCATTGCCGAAG 1027
```

```
CC   nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded
CC   on it, or a representative fragment or a sequence at least 95% identical
CC   to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC   to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus
CC   pneumoniae. The present invention also describes an isolated nucleic acid
CC   molecule encoding a homologue of any of the fragments of the S.pneumoniae
CC   genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC   by a process comprising: (a) screening a genomic DNA library using as a
CC   probe a target sequence defined by any of the sequences in SEQ ID NO:1 to
CC   391, identifying members of the library which contain sequences that
CC   hybridise to the target sequence and isolating the nucleic acid molecules
CC   from the members; or (b) isolating mRNA, DNA or cDNA produced from an
CC   organism, amplifying nucleic acid molecules whose nucleotide sequence is
CC   homologous to amplification primers derived from the fragment of the S.
CC   pneumoniae genome to prime the amplification and isolating the amplified
CC   sequences. The computer readable medium can be used in a computer-based
CC   system for identifying fragments of the S. pneumoniae genome of
CC   commercial importance, or expression modulating fragments of the S.
CC   pneumoniae genome. Products from the present invention can be used in
CC   diagnosis kits and assays, and pharmaceutical compositions and vaccines
CC   for S. pneumoniae
XX
SQ   Sequence 5910 BP; 1783 A; 1137 C; 1141 G; 1848 T; 0 U; 1 Other;

Query Match            19.1%;  Score 287.4;  DB 2;  Length 5910;
Best Local Similarity  51.7%;  Pred. No. 2.1e-58;
Matches  763;  Conservative  0;  Mismatches  696;  Indels  18;  Gaps  4;

Qy     11 AACGGCGCTTTTGGAATTTTGCTCCACCCCACCAGTTTTCCGGGTCGCTGGGGATTGGGG 70
          |          ||||||     |||  ||   |                   |||| |||
Db    108 AACGTCAAAGTGGTGTGTGATCGACATCTCTTCTCTTCCAGGAGCTTACGGAGTCGGAT 167

Qy     71 CTCTGGGCCGCGAGGCCGAGGCCGGTTTTGGACTGGCCGATGCGGAGCCGCCTGGT 130
          |      |||  ||||||        ||||  ||    ||   ||  ||  ||
Db    168 CATTTGGTCAAAGTGCTTACGACTTCGTTGATTTCTTGGTCCGTACAAAACACGTTACT 227

Qy    131 GGGCAGGTCTTACCGCTGGGCCCTACCAGTTACGGCGACTCGCCGTACCAGTTCCTTCTCGG 190
          ||  |||||||||    |||    |||||||||  ||     ||        |||| |||
Db    228 GGCAAATCCTTCCATTAGGAGCAACTAGTTACGGAGGATTCTCCTTACCAATCTTTCTCAG 287

Qy    191 CTTTTGCCGGTAACCCGTATTTGGTTGACCCCGAGATGCTGATTGAAAAAGGCTGGCTGG 250
          | ||| || | |||  ||||| |||||||  ||||||| ||| |      ||  |  ||
Db    288 CCTTCGCGGAAACACTCATTTTATCGATTTAGATATCTTGGTGGAGCAAGGTTTGTTGG 347

Qy    251 AACAAAGCGAAGCGCCCCCGGCATCCGACCCAGCGCGTGAATTATGGCGTGGCTTACC 310
          || ||||  ||||  ||  |||   ||||||||   |||  ||||||   || | ||
Db    348 AAGCAAGTGACCTTGAAGGAGTTGACTTTGGTAGCGATGCGTCGAAGTTGACTATGCTA 407

Qy    311 AGACCCGCTGGCCCGTTGCGCGGGGTTTCGCGGGACTTCGGGCAAGGGCTTCGGCCC 370
          | || |||||||  |  |||||||||  ||| |||  ||  ||| ||| ||  || |
Db    408 AAATCTACTATGCACGTCGTCCTCTTTTAGAAAAAGCCGGTGAAACGTTTCTTTGAAGTCG 467

Qy    371 AGGATAAGACCCGACTGAAGCCTTATCGAGGCGCGCTTCTGGGCTGGAAGACTATG 430
          ||||| || ||| |  |||   |||  ||  ||  |   |  |||| ||| ||||
Db    468 GAGAGTTAAAGATTTTGCTCAAGACAACAACCAATCATGGCGTTGAGCTCTTTG 527

Qy    431 CGCTCTTTATGGCCCTCAAGACCCGGTTGACCGGGAAGCCCTGGAACAGTGG--AGCC 487
          |  | |||||||  | ||||   |  ||||     ||| |  || || ||||  |
Db    528 CTGAGTATATGGCTATCAAAGAGTATTTTGACAATCTGTCTGGACTGAATGGCCAGATG 587

Qy    488 CCGAGCTGCGCGACCGTGAACCGGCTGCGCCCTGCGCCAGGGCCCGTGAGGAGCTGCGAGG 547
          |  ||||| | ||||||   |||||  |  | | || |||   |  ||| |||   |  |
Db    588 CAGATGCTCGGTCGTCGTAAAGCTTCAGCACTTGAAAGCTATCGTGAGCAATTGGCAGACA 647

Qy    548 AGGTGGCCGCTTTACGAGTGGATTCAGTGGCTTTTTATCTGGAATGGGGCCAGACCAAGG 607
          ||| ||    ||||| | |||| || || |  |||| |    ||||   | || | | ||
Db    648 AGTTGGTTTACCACCGTGACTCAATACTTCTTCCCAACATGGTTGAAACAATTGAAAG 707

Qy    608 CCTATGCCGAATCCAAGGGGATTCAAGATGGTATGCGGCGATATGCCCATCTTTGTCGGCCTTCG 667
          | |||||  |||| |||||| ||||| |||| ||||  | || |||||||  | |||| || |
Db    708 CTTACGCTAACGAACACCCACCACATGCGAAATCGTTGGGACATGCCAATCTACGTAGCCGGAAG 767

Qy    668 ATTCCTCAGATGTCTCGGCCACCCGCAGTACTTCTACCTCGAGCCGGATGGCAACCCCA 727
```

```
Db    968 GTTACAAGCTTTTGCAGCCGTTAAGGAAGAACTTGGTGAGCTAAACATCATCGCAGAAG 1027
          |      |||||| |  ||    ||    ||  ||||||||||| |||||||||||||| |
Qy   1028 ACCTGGGGTGATCACCCCGAGGTGGAGGCTTGCGCGATGGCTTCGGGTTCCCCGGCA 1087
          ||||  | ||||||||   ||||| |||| |  | |  | ||||   |||||| ||
Db   1028 ACCTTGGCTTCATGACAGATGAATTGCGTGAACGTACTGGCTTCCCAGGAA 1087
Qy   1088 TGAAGATTTGCAGTTTGCTTTTCCGGTGAGGACAACGCCTTTTGCCCCACAACTACC 1147
          ||||||||  |  ||||| ||||  |  |  ||   ||  ||||  ||    ||||
Db   1088 TGAAGATTCTTCAATTGCCTTCAACCCAGAAGACGAAAGCATTGATAGCCCACACT--- 1144
Qy   1148 CCGCGCACGGCAATGTGGTGGTGTACAGCGGACCCACGACACGACACCACCCTGGAT 1207
          |||  | |||  | |  ||  | ||| ||||  ||||||||   ||||||   ||| |||
Db   1145 TGGCACCTGCTAACTCAGTTATGTACACAGGACACACGATAACAATACACGGTTCTTGGTT 1204
Qy   1208 GGTTCCGCACCGCGCCGAGGCCGAGGCCGGCCCTTCATGCGGGCCTACCTGGCCCGCTATG 1267
          |||| ||  |  |   |  | |||  |  ||    ||||||| | |||| || |   ||||
Db   1205 GGTACCGTAATGAGATTGATGATGCG--------ACTCGGTGAGTACATGGCTCGTTACA 1255
Qy   1268 GCATCCGGTTGTTTGTCGGAATACGAGGTTCGCGGGCGCGCTTGATCGAGCTGGCCTTCAAAA 1327
          | || |||| || | |||  ||| |||   | |      || |||  | || |    ||| |
Db   1256 CGAACCGGTAAAGAATACGAAACAGTGGT---ACACGCCATGCTTCGTGTACAGTATTTTCAT 1312
Qy   1328 GCCCGGCCAAGCTGGCTATTGTGCCTTTGCAGGACGTGCGCTGGGGCTGGGCCCCGAGGCCC 1387
          |    |  | || ||| |||| ||  |||||| ||  |  |||| |  ||| |  ||| | |
Db   1313 CAGTTAGCTTTATGGCAATTGCAACTATGCAAGATTTACTAGAATTGGATGAGGCAGCTC 1372
Qy   1388 GCATGAACTTCCCCGGACGGCGGTGGGGGACAACTGGGCGTGGCGCTGGCCGAAGGGCGACC 1447
          | |||||||||| |  |  |  |||||| ||||||||  ||  | ||| ||||||||| |||
Db   1373 GTATGAACTTCCCATCTACCCCTTGGTGGGAAACTGGTCGTTCTTGGCGTATGCTGAAGATCAAT 1432
Qy   1448 TCGAGCCCGGTCTGGCGGGGACTGCGGGGGGCCCTGGC 1484
          | || || ||| || ||| |  |||     || ||| |
Db   1433 TGACACCAGCTGTGTCGAGGAAGGTTTGCTTGACTTGAC 1469
```

RESULT 9
AAV52298
```
ID   AAV52298 standard; DNA; 5910 BP.
XX
AC   AAV52298;
XX
DT   23-OCT-1998 (first entry)
XX
DE   Streptococcus pneumoniae genome fragment SEQ ID NO:165.
XX
KW   Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW   computer readable medium; vaccine; pharmaceutical composition; ds.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO9818931-A2.
XX
PD   07-MAY-1998.
XX
PF   30-OCT-1997; 97WO-US019588.
XX
PR   31-OCT-1996; 96US-0029960P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Kunsch CA, Choi GH, Dillon PJ, Rosen CA, Barash SC, Fannon M;
PI   Dougherty BA;
XX
DR   WPI; 1998-272225/24.
XX
PT   Computer-readable medium with recorded Streptococcus pneumoniae
PT   polynucleotide sequences - useful in diagnostic kits and assays, and
PT   pharmaceutical compositions and vaccines for Streptococcus pneumoniae.
XX
PS   Claim 1; Page 1063-1066; 1409pp; English.
XX
CC   The present invention describes a computer readable medium which has the
```

Tue Jan 17 09:28:39 2006

us-10-618-976-1.rng

```
WP    AB56454_12    1200001    1310000
WP    AB56454_13    1300001    1410000
WP    AB56454_14    1400001    1510000
WP    AB56454_15    1500001    1610000
WP    AB56454_16    1600001    1710000
WP    AB56454_17    1700001    1810000
WP    AB56454_18    1800001    1910000
WP    AB56454_19    1900001    2010000
WP    AB56454_20    2000001    2110000
WP    AB56454_21    2100001    2162598
```

```
Query Match          19.1%;     Score 287.4;  DB 10;  Length 110000;
Best Local Similarity  51.7%;   Pred. No. 4.3e-58;
Matches 763;  Conservative   0;  Mismatches 696;  Indels 18;  Gaps 4;
```

```
Qy    11  AACGGCGTTTTGGAATTTGCTCACCCACCAGTTTCCGGTGCGTCGGTGGGGGATTGGGG     70
              ||| |||||||||||||| |||| ||| ||| ||| ||||| || |||| || |||||||
Db 13861  AACGTCAAAGTGGTGTGTGATGCACATCTTCTTCTTCCAGGAGCTTACGGAGCTTACGGAT 13802

Qy    71  CTCTGGGCCCGAGGCCGAGCGGCGGTTTTGGACTGGCTCGTTGGTTCCGTGCCATGCGGAGCCCGCTCGT  130
              |||||||| |||| ||||| |||||||  |||  |  |||||| | |||||||||||
Db 13801  CATTTGGTCAAAGTGCTTCGGTGATTTCTGTAACATCGCAGAAAACAACGTTACT 13742

Qy   131  GGCAGGTCTTACCGCTGGGCCTGGGCGATGGCTTCGGGTTCCCGGCA  190
              |||| ||| || | || ||| || ||| ||||||||||||| ||||
Db 13741  GGCAAATCCTTCCATTAGGAGCAACTAGTTACGGGGAATTCTCCTTACCAAATCTTTCTCAG 13682

Qy   191  CTTTTTGCGGTAACCCGTATTGGTTGACCCGGAGGTGATTGAAAAAGGCTGGCTGG  250
              |||| ||||| || || ||||| ||| |||||
Db 13681  CCTTCGCAGGAAAACACTCATTTTATCGATTCGTTGGTGGACAAGGTTTGTTGTTGG 13622

Qy   251  AACAAAGCGAAGCGCCCGCGTATCCGACCCAGCGCGTGATTATGGCTGCTTTTACC  310
              || |||| |||||| |||| |||| || |||||| | |||| | ||||
Db 13621  AAGCAAGTGACCTTGAAGGAGTTGACTTTGGTAGCGGATGCCGTCTGAAGCTATGCTA 13562

Qy   311  AGACCCGCTGGCCCTGTTCGGGCCCCGGCCTTGGCGGGGTTTCGGGGCCTTAGTGAGGCGTTCGGGCCC  370
              |||||| | |||| || |||| |||| | ||||
Db 13561  AAATCTACTATGCACGTCGTCCTCTTTTAGAAAAAGCGGTGAAACGTTTCTTTGAAGTCG 13502

Qy   371  AGGATAAGACCCGACTGGAAGCCTTTATCGGAAGCCTGAGGC  430
              ||| || || |||| | |||||| ||| |||||||||
Db 13501  GAGATGTTAAAGATTTTGAGAATTTGCTCAAGAACAACCAATCATGGCTTCAGCTCTTTG 13442

Qy   431  CGCTCTTTATGGCCCTCAAGACCCGGTTGACGGCAAGCCCTGAACGAGTG---AGCC  487
              |||| |||||||| |||| ||| ||||| || |||||
Db 13441  CTGAGTATATGGCTATCAAAGAGTATTTTGACAATCTTGCTTGGACTGCCAGATG 13382

Qy   488  CCGAGCTGCGCAGACCGTGAACCGGTGGGCGCTGGCGCGTGGGCGGCTGCGCGAGG  547
              ||||| || || || |||||||||||||||
Db 13381  CAGATGCTCGTGCTCGTCGTAAAAGCTTCAGCAGTTGAAAGCTTGAAAGCTCAGACA 13322

Qy   548  AGTGGCCCTTTACGAGTGGATTCAGTGGCCTTTTTATCTCGGAATGGCC  607
              || ||||||||| ||||| |||| || |||
Db 13321  AGTTGGTTTACCACCGTGTGACTCAATACTTCTTCTTCCAACAATGGTTGAAATTGAAAG 13262

Qy   608  CCTATGCCGAATCCAAGGGAGATTCAGATTAATCGTTGGGGACAATGCCCAATCGTTGTGCCATCTTCG  667
              ||||| || || |||| || ||||| || |||||||
Db 13261  CTTACGTCAACGACAACCACCATCGTTGGGACATGCCAATCTCACGTAGCGGAAG 13202

Qy   668  ATTCCTCAGATGTCTGGGCCAACCCGCAGTACTTCTACCTCGGAGGCGATGGCAACCCCA  727
              |||| ||||||||||||| ||| || || ||| ||||||
Db 13201  ATTCAAGTGATATGTGGGCAAATCCACCATCTCTTCAAAACAGATGTCAATGGTCAATGTAAGGCTA 13142

Qy   728  CGGTGGTGCGGCGTTCGGGCGACTACTTCTCCGGGAACCGGCCAATC  787
              ||||| |||||||||||||| ||||| ||||||
Db 13141  CTTGTATCGCAGAATGCCCACCAGATGAGTTTCTGTAACTGGTAATC 13082

Qy   788  CGCTCATCGCTGGGATGTGATGGAAGGGACAACTTTGCCTGCGTGCATTGCCCGCATAA  847
              |||| ||||||||||
Db 13081  CAATCTTATGACTGGGAAGCAATGGACACCGGCTACAAAATGTGGATTGAACGCTTGC 13022

Qy   848  GGCAGTCGCTCAAGCAGTGCCACCTGGTCGCGCAATCGACCACCACTTCCGCCGGGTTTGAAGCCT  907
```

```
Db   768  ATTCAAGTGATATGTGGGCAAATCCAATCCATCTCTTCAAAACAGATGTCAATGGTAAGGCTA     827
Qy   728  CGGTGGTGCGGCGTTCCGCGGGGACTACTTCTCCGAAACCGGCCGCAGCTCTGGGGCAATC     787
Db   828  CTTGTATCGCGAGGATGCCCACCAGATGAGTTTTCTGTAACTGGTCAGCTTGGGGTAATC     887
Qy   788  CGCTCTATCGCTGGGATGTGATGGAAAGGGACAACTTTGCCTGCGTGCATTGCCCGCATAA     847
Db   888  CAATCTTATGACTGGGAAGCAATGGACCAAAGACGGCTACAAAATGTGGATTGAACGCTTGC     947
Qy   848  GGCAGTCGCTCAAGCAGTGCCACCTGGTCCGCATCGACCACCACTTCCGCGGTTTGAAGCCT     907
Db   948  GTGAAAGCTTCAAAATCTTACGATATCGTTCGTCCGTCCATCCGTGGCTTCCGGAATCTT    1007
Qy   908  ACTGGGAGGTTCCGTTTGGCCGGTGCGGGTGTGCGGTGCGGGGGCGCGTCTGGGTCAAGCCCCAG     967
Db  1008  ACTGGGAAATCCCTGCTGGTTCCGATACAGCAGCACCTGGTGAGTGGGTGAAAGGGTCCAG    1067
Qy   968  GGGAGAAGCTGTTTGCTGCGGTGCGGGCCCAACTGAGCGATGCGCCCATCATTGCCGAAG    1027
Db  1068  GTTACAAGCTTTTTGCAGCCGTTAAGGAGAACTTGGTGAGCGAGCTAAACATCACGCAGAAG    1127
Qy  1028  ACCTGGGGGTGATCACCCCCGAGGTGGAGGCTTTGCCGCGATGGCTTCGGGTTCCCGGCA    1087
Db  1128  ACCTTGCTTCATGACAGATGAAGTGATCGAATTGCGTGAACGTAGCCTGGCTTCCCAGGAA    1187
Qy  1088  TGAAGATTTTGCAGTTTGCTTTTCCGGTGAGGACAACGCCCTTTTGCCCCACAACTACC    1147
Db  1188  TGAAGATTTCTCAATTGCCTTCAACCCCAGAAGCATTGATAAGCCACCACT---       1244
Qy  1148  CCGCGCACGGCAATGTGGCTGGTGTACAGCGGAACGGAACACCCTGGGAT           1207
Db  1245  TGGCACCTGCTAACTCAGTTATGTTACACAGGAACACAATAACATACGGTTCTTGGTT    1304
Qy  1208  GGTTCCGACCGCACCGCGCCGGAGCGGCCTTCACTACCTGGCCCGTATG          1267
Db  1305  GGTACCCGTAATGAGATTGAATGGATGATGCG------ACTCGTGAGTACATGGCTCGTTACA  1355
Qy  1268  GCATCCGTGTTGTTGTGTTCGGAATACGAGAGTGTCGGCGGCCTTGCGGATGAGCGTGAGCGTGAG  1327
Db  1356  CGAACCGTAAAGAATACGAAACAGTGT---ACACGCTATGCTTCGTACACAGTATTTTCAT  1412
Qy  1328  GCCCGGCCAAGCTGGCTATTGTGCCTTTGCAGGACGTGCTGGGGCGTGGGCCCCCGAGGCCC  1387
Db  1413  CAGTTAGCTTTATGGCAATGCAACTATGCAAGATTTCAAGAATTGGATGAATTGGATGAAT  1472
Qy  1388  GCATGAACTTCCCCGACGGCGTGGGGACAACGGCCTGGCGCTGCACGCCGAAGGCGACC    1447
Db  1473  GTATGAACTTCCCATCTACCCCTTGGTGAAACTGGTCTTTGGGCGTATGCTGAAGATCAAT  1532
Qy  1448  TCGAGCCGGGTCTGGCCCGGCGGACTGCGGGGCCCTGGC    1484
Qy  1533  TGACACCGAGCTGTCGAGGAAGGTTTGCTTGACTTGAC    1569
```

```
RESULT 10
AB56454_19/c
Continuation (20 of 22) of AB56454 from base 1900001 (Streptococcus pneumoniae type 4
WP Sequence split into 22 fragments    LOCUS AB56454 Accession AB56454
WP    Fragment Name    Begin        End
WP    AB56454_00       1            110000
WP    AB56454_01       100001       210000
WP    AB56454_02       200001       310000
WP    AB56454_03       300001       410000
WP    AB56454_04       400001       510000
WP    AB56454_05       500001       610000
WP    AB56454_06       600001       710000
WP    AB56454_07       700001       810000
WP    AB56454_08       800001       910000
WP    AB56454_09       900001       1010000
WP    AB56454_10       1000001      1110000
WP    AB56454_11       1100001      1210000
```

XX
PT New isolated nucleic acid encoding a Streptococcus pneumoniae
PT polypeptide, useful for diagnosing, preventing and/or treating
PT pathological conditions resulting from the bacterial infection.
XX
PS Disclosure; SEQ ID NO 336; 151pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising a sequence
CC encoding a Streptococcus pneumoniae polypeptide ADR91366polypeptide, or its
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,
CC ADR94489, ADR94800, ADR94837, ADR94969, ADR95251, ADR95642, ADR95682,
CC ADR96079) or any of the fully defined sequences appearing as ADR91705,
CC ADR91886, ADR91917, ADR92197, ADR92234, ADR93033, ADR93079, ADR92366, ADR92650 or
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide
CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is
CC hybridisable under high stringency conditions to the nucleotide sequence.
CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.
CC Also included are a recombinant expression vector comprising the isolated
CC nucleic acid cited above operably linked to a transcription regulatory
CC element, a cell comprising the recombinant expression vector and a probe
CC comprising at least 20 consecutive nucleotides of the nucleotide
CC sequences as cited above. The methods and compositions of the present
CC invention are useful for the diagnosis, prevention and/or treatment of
CC pathological conditions resulting from bacterial infection by
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and
CC otitis media. The present sequence is one of the 2603 disclosed s.
CC pneumoniae nucleic acid sequences. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=6800744B1.
XX
SQ Sequence 1377 BP; 398 A; 277 C; 323 G; 379 T; 0 U; 0 Other;

Query Match   17.7%; Score 265.6; DB 13; Length 1377;
Best Local Similarity   52.1%; Pred. No. 2.4e-53;
Matches 698; Conservative   0; Mismatches 624; Indels 18; Gaps 4;

Qy 148 GGCCTACCAGTTACGGCGACTCGCCGTACCAGTCCTTCTCGCTTTTGCCGGTAACCCG 207
          | ||||| |||||||||||| | || | |||| || ||||||| |||||||
Db   4 GGAGCAACTAGTTACGGCGGGGATTCTTCCTTACCAATCTTTCTCAGCTTCCCAGGAAACACT 63

Qy 208 TATTTGGTTGACCCCGAGATGCTGATTGAAAAAGGCTGCTCGACAAAGCAAAGCGAAGCGGCCC 267
          | ||| || | | ||| | || |||||| || || | ||| | ||
Db   64 CATTTTATCGATTTAGATATCTTGGTGTGGAGCAAGGTTTGTTGGAAGCAAGTGACCTTGAA 123

Qy 268 CCGCCGTATCCGACCCAGCGCGTGGAGTTATGGCGGATTATGGCGCCTGGCCCCTG 327
          || | ||| | ||||||||| | ||| || | | |||| ||||
Db 124 GGAGTTGACTTTGGTAGCGATGCGTCGAAGTTGACTATCGTGAAGTTGACTACTATGCACGT 183

Qy 328 TTGCGGCGGGGCTTTCGGCGGGGTTTCGGCGCAAGGGCTTCGGGCAAGCGTTCTTTGAAGTCGGA 387
          ||||| ||| || || ||||| | || || ||| ||||| |||| ||| |||||| ||
Db 184 CGTCCTCTTTAGAAAAAGCGGTGAAACGTTCTTTGAAGTCGGAGATGTTAAAGATTT 243

Qy 388 GAAGCCTTTATCGAGCGGCCGAGCCGTTCTGGAAGACTATGCGCTCTTTATGGCCCTC 447
          |||||| || | || || |||| | ||||| | | || || || | |||
Db 244 GAGAAATTTGCTCAAGACAACCAATCATGGCTTCAAGGCCAGATCGAATGGCTATC 303

Qy 448 AAGACCCGGTTTGACGGCAAGCCCTGGAACGAGTGGAGC---CCCGAGCTGCCGCGACCGT 504
          |||| | ||| || || | | | || || | | ||| ||| || ||| | |||
Db 304 AAAGAGCATTTTGACAATCTTGCTTGGACTGGAACGAATGGCCAGATGCAGATGCTCGTGCTCGT 363

Qy 505 GAACCGGCTGCCCTGGCCCAGGGCCCGTGAGGAGCTCGGACCCTTTACGAG 564
          || |||| | || | ||| | | | || | || | || || ||||| |
Db 364 AAAGCTTCAGCACTTGAAAGTATCGTGAAGCATCGTGAGCAATTGGCAGACAAGTTGGTTTACCACCGT 423

Qy 565 TGGATTCAGTGGCTTTTTATCTGGAATGGGCCCAGACCGGCCTATGCGGAATCCAAG 624
          || |||| ||| | | |||| || | | | || ||| | |||
Db 424 GTGACTCCAATACTTCTTCTCCAACATGTGTTGAAATTGAAAGTTACGTTACGGACACAAC 483

Qy 625 GGGATTCAGATTATCGGCGAATATGCCCATCTTTGTGGCGCTTCGATTCCTCAGATGTCTGG 684
          | ||| | | | || | | || || | | |||| | || |||| |||||| ||
Db 484 CACATCGAAATCGTTGGGGACACTAGCTACGTAGCGGAAGATTCAAGTGAAGATATGTGG 543

RESULT 11
ADR91701
ID ADR91701 standard; DNA; 1377 BP.
XX
AC ADR91701;
XX
DT 16-DEC-2004 (first entry)
XX
DE Novel S. pneumoniae DNA sequence, SEQ ID 336.
XX
KW Meningitis; bacteraemia; pneumonia; otitis media; ds;
KW bacterial infection.
XX
OS Streptococcus pneumoniae.
XX
XX
PN US6800744-B1.
PD 05-OCT-2004.
XX
PF 30-JUN-1998; 98US-00107433.
XX
PR 02-JUL-1997; 97US-0051553P.
PR 12-MAY-1998; 98US-0085131P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
DR WPI; 2004-697205/68.
DR P-PSDB; ADR94304.

Db 13021 GTGAAAGCTTCAAAATCTACGATATCGTTCGTATCGACCACTTCGGTGCTTCGAATCTT 12962
Qy   908 ACTGGGGAGGTTCCGTTCGCCGGCCCAATGCTGTGGAGGGGCCTGGGCTCAAGCCCCAG 967
Db 12961 ACTGGGAATCCCTGCTGGTTCCGATACAGCAGCACCTGGTGAGTGGGGTGAAAGGTCCAG 12902

Qy   968 GGGAGAAGCAGCTGTTGTCGCGCGGCGGGCCCAACTGAGCCGATGCGGCCCATCATTGCCGAAG 1027
Db 12901 GTTACAAGCTTTTTGCAGCCGTTAAGGAAGAACTTGGTGAGCTGAGCTAAACATCATCGCAGAAG 12842

Qy 1028 ACCTGGGGGTGATCACCCCGAGGTGAGAGGCTTTGCGCGATGCTTCGGTTCCCCCGGCA 1087
Db 12841 ACCTTGGCTTCATGACAGATGAGAGTGAATTGCCGTGAACGTACTGCTTCCCAGAA 12782

Qy 1088 TGAAGATTTGCAGTTTGCTTTCCGTGAGGACAACGGCCTTTTGCCCCAACTACC 1147
Db 12781 TGAAGATTCTTCAATTGCCTTCAACCCAGAACGAAAGCATTGATAGCCACACT--- 12725

Qy 1148 CCGCGACGGCAATGTGGTGTCACGCGAACCCAGCACACACGACACCACCCTGGGAT 1207
Db 12724 TGGCACCTGCTAACTCAGTTAGCTATGTAGCACAGGAACACGATAACAATACGGTTCTTGGTT 12665

Qy 1208 GGTTCCGCACCGCGCCGAGGCCGAGCGGCCCTTCATGCGCGGCCTACCTGGCCCGCTATG 1267
Db 12664 GGTACCGTAATGAGATTGATGATGCG---------ACTCGTGAGTACATGGCGTCGTTACA 12614

Qy 1268 GCATCCGTTGTTGTTCTCGGAATACGAGTCGGGGCGCTTTGATCGAGCTGGGCCTTCAAAA 1327
Db 12613 GGTACCGTAAAGAAGAGTAAAGAACAGTGGT---ACACGCTATGCTTCGTACAGTATTTTCAT 12557

Qy 1328 GCCCGGCCAAGCTGGCTATTGTGCCTTTGCAGGACGTGCTGGGGCGTGCTGGGCCCCGAGGGCC 1387
Db 12556 CAGTTAGCTTTATGGCAATTGCAACTATGCGCAACTGGCAAGATTGGGAATTGGATGGAGGCAGCTC 12497

Qy 1388 GCATGAACTTCCCCGGACGGCTGGGGGACAACCTGGGCGCTGGGCGCTACGCCGAAGGGCGACC 1447
Db 12496 GTATGAGACTTCCCCATCTACCCTTGGTGAAACTGGGTCTTGGCTATGACTGAAGATCAAT 12437

Qy 1448 TCGAGGCCCGGTCTGGCCGGCCGGGCGGAGGGGCCCTGGC 1484
Db 12436 TGACACCAGCTGTGTCGAGGAAGTTTGTTGCTTGAC 12400

```
XX
PD    23-JUN-2005.
XX
PF    10-JUL-2003; 2003US-00617320.
XX
PR    02-JUL-1997;   97US-0051553P.
PR    12-MAY-1998;   98US-008513lP.
PR    30-JUN-1998;   98US-0010743.
XX
PA    (DOUC/) DOUCETTE-STAMM L A.
PA    (BUSH/) BUSH D.
XX
PI    Doucette-Stamm LA, Bush D;
XX
DR    WPI; 2005-477576/48.
DR    P-PSDB; AEA58174.
XX
PT    New isolated nucleic acid molecules and encoded polypeptides useful for
PT    diagnosing, preventing or treating bacterial infections, particularly
PT    Streptococcus pneumoniae infection.
XX
PS    Claim 1; SEQ ID NO 336; 144pp; English.
XX
CC    The invention relates to an isolated nucleic acid molecule for detecting,
CC    preventing or treating pathological conditions resulting from bacterial
CC    infection. The isolated nucleic acid comprises: (a) any of the 2603
CC    nucleotide sequences of AEA55236 to AEA57838; (b) a nucleotide sequence
CC    encoding a Streptococcus pneumoniae polypeptide comprising any of the
CC    2603 amino acid sequences of AEA57839 to AEA60441; or (c) a nucleotide
CC    sequence of at least 8 nucleotides in length, where the sequence is
CC    hybridizable to a nucleic acid having any of the nucleotide sequences in
CC    (a). Also described: (1) a recombinant expression vector comprising the
CC    above nucleic acid operably linked to a transcription regulatory element;
CC    (2) a cell comprising the recombinant expression vector; (3) producing an
CC    S. pneumoniae polypeptide; (4) a probe comprising a nucleotide sequence
CC    consisting of at least 8 nucleotides of any of AEA55236 to AEA57838; (5)
CC    treating a subject for S. pneumoniae infection; (6) a recombinant or
CC    substantially pure preparation of an S. pneumoniae polypeptide or its
CC    fragment, where the polypeptide is selected from AEA57839 to AEA60441;
CC    (7) a vaccine composition for preventing or treating an S. pneumoniae
CC    infection, comprising an amount of the above nucleic acid or polypeptide;
CC    (8) detecting the presence of a Streptococcus nucleic acid in a sample;
CC    (9) a computer readable medium having recorded the nucleotide sequences
CC    of AEA55236 to AEA57838; (10) a computer based system for identifying
CC    fragments of the Streptococcus genome of commercial importance. The
CC    composition and methods are useful for diagnosing, preventing or treating
CC    bacterial infections, particularly S. pneumoniae infection. The present
CC    sequence represents a S. pneumoniae ORF nucleic acid sequence from the
CC    present invention. Note - The sequence data for this patent did not form
CC    part of the printed specification, but was obtained in electronic format
CC    directly from the USPTO web site.
XX
SQ    Sequence 1377 BP; 398 A; 277 C; 323 G; 379 T; 0 U; 0 Other;

Query Match             17.7%;  Score 265.6;  DB 14;  Length 1377;
Best Local Similarity   52.1%;  Pred. No. 2.4e-53;
Matches   698; Conservative    0; Mismatches   624; Indels  18; Gaps    4;
```

```
Qy        148 GGCCCTACCAGTTACGGCGACTCGCCGTACCAGTCTTCTCGGCTTTTGCCGGTAACCCG  207
                ||| || || |   ||    |||  ||       |||    |||
Db          4 GGAGCAACTAGTTACGGCGGGATTCCCTTACCAATCTTTCTCAGCTTCGCCAGGAAACACT  63

Qy        208 TATTTGGTTGACCCCGAGATGCTGATTGAATGAAAAAGGCTGGCTGAAACAAAGCGAAGCGGCCC  267
                ||  | | |||| |||  ||  ||||| | |||| |||   |||  ||  || ||
Db         64 CATTTTATCGATTTAGATATCTTGGTGGAGCAAGGTTTGTTGGAGCAAGTGACCTTGAA  123

Qy        268 CCGCCGTATCCGACCCAGCGCGTGGATTATGGCGTGATTATGGCCTGGCTTTACCAGCACCCCGTCCCCCTG  327
                |||| ||  |||| ||| ||  || | |  |||||||| |||    || || ||  ||
Db        124 GGAGTTGACTTTGGTAGCGATGGCGTCGAAGTTGACTTGGCTAAATCTACTACTGCACGT  183

Qy        328 TTGCGGCGGGCTTTCGGCGGGTTTCGGGCAACGTTTCTTGAAGTCGGAGATGTTAAAGATTTT  387
              |||  || |||  ||  |  |  |    |||| ||| |||| ||   ||| ||  ||| |
Db        184 CGTCCTCTTTAGAAAAAGGCGGTGAAACGTTTCTTGAAGTCGGAGATGTCGGAGATGTT  243
```

```
Qy        685 GCCAACCCGCAGTACTTCTACCTCGAGGCGGATGGCAACCCCACGTGGTGGGCGGCGTT  744
                ||| || |                |||||
Db        544 GCAAATCCACATCCTTCAAAACAGATGTCAATGGTAGGCTACTTGGTATCGCAGGATGC  603

Qy        745 CCCGCGGGACTACTTCTCCGAAACCGGCCAGCTCTGGGGCAATCCGCTCTATCGCTGGGAT  804
                  |||||  || || ||  ||| ||  |||||||
Db        604 CCACCAGATGAGTTTTTCTGTAACTGGTCAGCTTTGGGGTAATCCAATCTATGACTGGGAA  663

Qy        805 GTGATGGAAAGGGACAACTTGCCTGGTGCATTGCCCGCGATAAGGCAGTCGCTCAAGCAG  864
                |  || |||  || || ||||| || ||||    ||| ||
Db        664 GCAATGGACAAAGACGGCTACAAATGGTCGATTGAACGCTTGCCGTGAAAGCTTCAAAAATC  723

Qy        865 TGCCACCTGGTGGCGCATGCACACTTCCGCGGCTTGAAGCCTACTGGGAGGGTTCCGTT  924
                 || ||    ||| |   | ||| || ||  |   |||| |
Db        724 TACGATATCGTTCGTATCGACCACTTCCGTGGCTTCGAATCTTACTGGGAAATCCCTGCT  783

Qy        925 GGCCGGCCCAATGCTGTCGGAGGGGCGCTGGGTGGGTCAAAGCCCCAGGGGAGAAGCGTGTTTGCT  984
                ||  || ||  | ||  || || | | | ||
Db        784 GGTTCCGATACAGCAGCACCTGGTGAGTGGTGAGTGGGTGAAAGGTCCAGGCTACAAGCTTTTTGCA  843

Qy        985 GCGGTGCGGGCCCAACTGAGCCGATGCGCCTAAACATCGCAGAAGACCTTGGCTTCATGACA  1044
                |||  ||| |||  || || |  ||  | ||  |||| ||| ||||| | || ||
Db        844 GCCGTTAAGGAAGAACTTGTGAGCTAAACATCATCGCAGAAGACCTTGGCTTCATGACA  903

Qy       1045 CCCGAGGTGGAGGCTTGCGGCGATGGCGTTCCCCGGTCTCGGTGCCCGCATGAAGAGATTTTGCAGTTT  1104
                |||||  ||  |||  ||| ||| ||| ||    |||| || |   |||| ||  ||
Db        904 GATGAAGTGATCGAATTGCGTGAACGTACTGGCTTCCCAGGATTGAAGATTCTTCAATTT  963

Qy       1105 GCTTTTTCCGGTGAGGACAACGGCCTTTTGCCCCACAACTACCCCGCGGCAATGTG  1164
                ||  |    || |    ||  ||  |     |||||   ||    |||  |  || |
Db        964 GCCTTCAACCCAGAAGACGAAAGCGATTGATAGCCCACACT---TGGCACCTGCTAACTCA  1020

Qy       1165 GTGGTGTACAGCGGAACCCACGACGAACGACCACCACCCCTGGGATGGTTCCGCACCGCGCCG  1224
                |||  ||| |||  ||  | || || | | | |  |||  ||| ||
Db       1021 GTTATGTACACAGGAACACAGCAGACAATAACGATTACGGTTCTTGGTACCGTAATGAGATT  1080

Qy       1225 GAGGCCGAGCGGCGGGCCTTCATGCGGGCCTACCTGGCATGCGACATCCGTGTTGTTGTCG  1284
                | |    |    |||                   ||||  ||   ||| |||| |  ||
Db       1081 GATGATGCG------ACTCGTGAGTACATGGCTCGTCGTACACGAACCGTAAAGAGAATAC  1131

Qy       1285 GAATACGAGGTCGCGGGCCGTTTGATCGAGCTGCGCCTTCAAAAGCCCGGCCAAGTCGGCT  1344
                |||  ||| |||  ||   || ||  |||| |  ||   ||||| |||  || || ||
Db       1132 GAAACAGTGGT---ACACGCTATGGCTTCTCGTACACAGTATTTTCATCAGTTAGCTTTATGGCA  1188

Qy       1345 ATTGTGCCTTTGCAGGACGTGCTGGGGCTGGGGCCCCGAGGCCCGCCATGAACTTCCCCGGA  1404
                |||  || | ||  | ||  |  |||| ||  | |||  | ||    ||| ||||  |||
Db       1189 ATTGCAACTATGCAAGATTTACTAGAATTGGATGAGGCAGCTCGTATGAACTTCCCATCT  1248

Qy       1405 CGGCTGGGGGACAACTGGGTCTTGGCGTACGCCGAAGGGACCTCGAGCCCGTCTCGGCC  1464
                |||  ||||| | |  ||  ||  |   | |||   |||   | || || || |  ||
Db       1249 ACCCTTGGTGGAAACTGGTCTTGGCGTATGACTGAAGATCAATTGACACACCAGTTGTCGAG  1308

Qy       1465 GCGGGACTGCGGCGCCCTGGC  1484
                |   |||| |||  ||  |
Db       1309 GAAGGTTTGCTTGCTTGAC  1328
```

```
RESULT 12
AEA55571
ID   AEA55571 standard; DNA; 1377 BP.
XX
AC   AEA55571;
XX
DT   25-AUG-2005  (first entry)
XX
DE   Streptococcus pneumoniae ORF nucleic acid sequence SEQ ID NO:336.
XX
KW   bacterial infection; Streptococcus pneumoniae infection; antibacterial;
KW   vaccine; gene; ds.
XX
OS   Streptococcus pneumoniae.
XX
PN   US2005136404-A1.
```

Qy   1465 GCGGGACTGCGGCGCCCTGGC 1484
            ||| || || ||||
Db   1309 GAAGGTTGCTTGACTTGAC 1328

RESULT 13
ADQ15021
ID   ADQ15021 standard; DNA; 1731 BP.
XX
AC   ADQ15021;
XX
DT   23-SEP-2004 (first entry)
XX
DE   DNA encoding potato amylo maltase polypeptide.
XX
KW   Amylo maltase; transgenic; foodstuff; food additive; modifier;
KW   saccharide; (alpha)-1,4-glucan; cooked rice; Japanese confectionery;
KW   snack; noodle; Chinese dumpling; shao-mai; fishery kneading;
KW   refrigerated; baby food; pet; animal feed; drink; food supplement;
KW   cyclic glucan; potato; gene; ds.
XX
OS   Solanum tuberosum.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..1731
FT   /tag= a
FT   /product= "Potato amylo maltase polypeptide"
XX
PN   JP2004187674-A.
XX
PD   08-JUL-2004.
XX
PF   12-SEP-2003; 2003JP-00322319.
XX
PR   26-NOV-2002; 2002JP-00342966.
XX
PA   (EZAK) EZAKI GLICO CO LTD.
PA   (NIDE) NEC CORP.
XX
DR   WPI; 2004-528719/51.
DR   P-PSDB; ADQ15022.
XX
PT   New mutated amylo maltase polypeptide having increased enzyme activity
PT   and reduced hydrolyzing activity, useful for manufacturing foodstuffs
PT   such as Japanese confectionery, noodles, baby foods, and food additives.
XX
PS   Example 9; SEQ ID NO 12; 51pp; Japanese.
XX
CC   The invention relates to a novel amylo maltase polypeptide comprising the
CC   amino acid sequence of a wild-type amylo maltase with a substitution,
CC   addition or deletion at a position, where the amino acid residue
CC   interacts with acarbose of amino acids other than the wild-type amylo
CC   maltase amino acid sequence. The invention further comprises: a nucleic
CC   acid molecule containing a nucleic acid sequence encoding the novel amylo
CC   maltase polypeptide; a vector containing the nucleic acid sequence; a
CC   cell containing the nucleic acid sequence; biological tissue containing
CC   the nucleic acid sequence; a transgenic organism containing the nucleic
CC   acid sequence; a foodstuff, a food additive or a modifier of foodstuff
CC   containing the nucleic acid sequence; a computer readable recording
CC   medium which contains the information of the nucleic acid sequence
CC   encoding the amino acid sequence of the novel amylo maltase polypeptide;
CC   and a cyclic glucan obtained by reacting the novel amylo maltase
CC   polypeptide on saccharides which have a linear structure of (alpha)-1,4-
CC   glucan. The novel amylo maltase polypeptide is useful for manufacturing
CC   foodstuffs such as cooked rice, Japanese confectionery, snacks, bakeries,
CC   noodles, Chinese dumpling, shao-mai, fishery kneading goods,
CC   refrigerating foods, baby foods, pet foods, feed for animals, drinks, and
CC   food supplements, where the method involves adding the novel amylo
CC   maltase polypeptide to the foodstuff before or immediately after heat
CC   processing of the raw material, where the novel amylo maltase polypeptide
CC   generates a cyclic glucan from the starch of the foodstuff. The novel
CC   amylo maltase polypeptide is useful for manufacturing glucan which has a
CC   cyclic structure of (alpha)-1,4-glycoside linkage, foodstuff materials

Qy   388 GAAGCCTTTATCGAGGCCGAGCGCTTCTGGCTGGAAGACTATGCGGTCTTTTATGGCCCTC 447
           |||||||| ||| || | || ||| ||||
Db   244 GAGAAATTTGCTCAAGACAACCAATCATGGCTTGAGCTCTTGCTGAGTATATGGCTATC 303

Qy   448 AAGACCCGGTTTGACGGCCAAGCCCTGGAACGAGTGGAGC---CCCGAGCTGCGCGACCGT 504
           || ||||||| ||||||| | |||
Db   304 AAAGAGCATTTTGACAATCTTGCTTCGGACTGAATGGCCAGATGCCAGAGGCTTCGTGCTCGT 363

Qy   505 GAACCCGCTGCCCTGGCCAGGGCCCGTGAGGAGCTGGCCCTTTACGAG 564
           | |||| || | || | | ||||
Db   364 AAAGCTTCAGCACTTGAAAGCTATCGTGAGCAATTGGCAGACAAGTTGGTTTACCACCGT 423

Qy   565 TGGATTCAGTGGCTTTTTTATCTGGAATGGGGCCAGACCAAGGCCTATGCCGAATCCAAG 624
           |||| ||||| | || | || | | | |||
Db   424 GTGACTCAATACTTCTTCTTCCAACAATGGTTGAAAATTGAAAGCTTACGCTAACGACAAC 483

Qy   625 GGGATTCAGATTATCGGCGATATGCCCATCTTTGTGGCCTTCGATTCCTCAGATGTCTGG 684
           ||| ||| ||| | | | | ||| || | || || ||
Db   484 CACATCGAAATCGTTGGGGACATGCCAATCTACGTAGCGGAAGATTCAAGTGATATGTGG 543

Qy   685 GCCAACCCGCAGTACTTCTACCTCGAGGCCGATGGCAACCCCACGCTGGTGTGGCGGGCGTT 744
           ||| | || | || | | ||||| || |||
Db   544 GCAAATCCAATCTCTCTCAAAACAGATGGTAAGGCTACTTGTATCGCAGGATGC 603

Qy   745 CCGCGGGACTACTTCTCCGAAACCCGGCCAGCTCTGGGGCAATCCGTCTCTATCGCTGGGAT 804
           |||| || | |||| ||| ||||| | | | | || ||||
Db   604 CCACCAGATGAGTTTTCTGTAACTGTCAGCTTTGGGTTTGGGTAATCCAATCTATGACTGGGAA 663

Qy   805 GTGATGGAAACGGGACAACTTTGCCCTGGTGCATTGCCCGGCCAGTAAGGCCAGTCAAGCAG 864
           ||||| ||| |||| | || | | || |||| | | | ||||| ||
Db   664 GCAATGGGACAAAGACGGCTACAAATGGTGACACTTCCGGGGGTTTGAAGCGCTTGGGAGGTTCCGTTT 723

Qy   865 TGCCACCTGGTCGCGATCGACCACTTCCGCGGGGTTTGAAGCGTTTGGGAGGTTCCGTTT 924
           | |||| ||| || | ||| ||| | || | ||||| | |||
Db   724 TACGATATCGTTCGTATCGACCACTTCCGCTGTGCTTCGAATCTTACTGGGAAATCCCTGCT 783

Qy   925 GCCGGCCCAATGCTGTGAGGAGGGGCGCTGGCTCAAAGCCCAGGGGAGAAGCTGTTTTGCT 984
           | ||| |||||| ||||| | || ||| || |||| || || | | |
Db   784 GGTTCCGATACAGCAGCACCTGGTGAGTGGTGAGGTGAAGGTCCAGGCTACACAGCTTTTGCA 843

Qy   985 GCGGTGCGGGCCCAACTGAGCGATGCGCCCATCATTGCCGAAGACTGGGGGGTGATCACC 1044
           || ||| ||| | || || ||| | || ||||| | || ||
Db   844 GCCGTTAAGGAAGAACTTGGTGAGCTAAACATCATCGCGAAGAACCTTGGCTTCATGACA 903

Qy   1045 CCCGAGGTGAGGGTTTGCGGCGATAGCGATCGCTTGTTCCCCGGCATGAAGATTTGCAGTTT 1104
            || | |||||| ||| || || | | | | || || ||||| || ||||
Db   904 GATGAGTGATCGAATGCCGATAGCTGAACGTACTGGCTTCCGGTTCTTCCCAGAATGAGAGATTCTTCAATTT 963

Qy   1105 GCTTTTTCCGGTGAGGACAACGCCCTTTTGCCCCACAACTACCCGGCGCACGGGCAATGTG 1164
            ||||| ||| | |||||| | | || ||| || || || | | ||||||
Db   964 GCCTTCAACCCAGAGACGGAAAGCATTGATGATGCCCACACT---TGCACCTGCTAACTCA 1020

Qy   1165 GTGGTGTACAGCGGGAACCCACGACAACGACACCACCGCGCCG 1224
            || | || ||| | ||| || || |||| |||
Db   1021 GTTATGTACACAGGAACACACGATAACATACGGTTCTTCGTTGGTACCGGTAATGAGATT 1080

Qy   1225 GAGGCCGGAGCGGGCCTTCATGCTACGGGCCGCTACCTGCCGCATGGCGACATCCGTTGTTGTCG 1284
            ||| ||||| |||| | || || | || | || | || | |
Db   1081 GATGATGCG--------ACTCGTGAGTACATACGTGGGCTGGTCGTTACCGGAACCGTAAAGAATAC 1131

Qy   1285 GAATACGAGGTGCGCGGGCGCCTTTGATCGAGCTGGCCTTCAAAAGCCCGGCCAAGCTGGCT 1344
            || ||| || | | | | || | || |||| ||
Db   1132 GAACAGTTGT---ACACGGCTATGTCTTCGTACAGTATTTTCATCAGTTAGCTTTTATGGCA 1188

Qy   1345 ATTGTGCCTTTGCAGGACGTGCTGGGGCTGGGGCGCTGGGGGCCCCGAGGCCCCATGAACTTCCCCGGA 1404
            ||| | || | |||| || || | | |||| | | | | |||||
Db   1189 ATTGCAACTATGCAAGATTTACTAGAATTGCATGAGGGCAGCAGTCGTATGAACTTCCCATCT 1248

Qy   1405 CGGCTGGGGGACAACTGGTCTTGGCGTGGCGTGCTACGCCGAAGGGCCACCTCGCAGCTGTCGGCC 1464
            ||| ||||| || | ||| | | | || | | | | | | |||| || | |||
Db   1249 ACCCTTGGTTGGAAACTGGTCTTGGCGTATGACTGAAGATCAATTGACACCCAGCCGTGTCGGCGAG 1308

CC   and food additives, modifier of foodstuffs, a drink or eatable
CC   composition, infusion solution or composition for adhesion, where the
CC   method involves reacting the saccharide containing the linear structure
CC   of (alpha)-1,4-glycoside linkage with the novel amylo
CC   maltase polypeptide. This polynucleotide sequence represents the DNA
CC   encoding a potato amylo maltase polypeptide of the invention.
XX
SQ   Sequence 1731 BP; 485 A; 344 C; 405 G; 497 T; 0 U; 0 Other;

Query Match   12.1%;   Score 181.4;   DB 12;   Length 1731;
Best Local Similarity 48.3%;   Pred. No. 3.3e-33;
Matches 689;   Conservative 0;   Mismatches 701;   Indels 37;   Gaps 5;

```
Qy        21 TGGAATTTGCTCCACCCACCAGTTTCCGGGTCGCTGGGGGATTCGGGGCTCTGGGCCG        80
             |||||||||||||  ||||  ||  | |||| |     ||  |||| |||||||
Db       240 TGGAATTTGCTCCATCCGACGTCGTTTCTGGACCTTATGGTATTGGTGACCTTGCCCC       299

Qy        81 CGAGGGCCGAGCGGTTTTTGGACTGGCTGGCGGATGCGGGAGCCCGTCGTGGTGCAGGTCTT      140
             |||||||||    ||| |||| |  ||  |||||||||   |||||| |||||||
Db       300 TCAGGGCTTTAAGTTCCTGATTGGGCTTCATCTTGCTGGTTGTCTCCCTTTGCGCAGGTTCT      359

Qy       141 ACCGGTCTGGGCCCTAC--------CAGTTACGGCGACTCGCGTACCAGTCCTTCTC      188
             |||||||            ||      ||   ||||||| ||
Db       360 TCCACTTGTACCGCCTGGAAAGAGAGGCAATGAAGATGGATCACCCTATTCAGGACAGGA      419

Qy       189 GGCTTTTGCCGGTAACCCGTATTTGGTTGACCCCGAGATGCTGGATTGAAAAAGGCTGGCT      248
             ||||||          |||||||| |||| |||||||| |   || || || ||||
Db       420 TGCAAATTGTGAAACACACTTCTGATTTCTCTGAAGAGCTTGTTGATGATGGTTTACT       479

Qy       249 GGAACAAAGCGAAGCGCCCCCCGTATCCGACCCAGCGCGTGATTATGGCTGGCTTTA      308
             ||          ||||| || ||||  |    |  |||||||||||| ||
Db       480 GAAGATGGAGGAGCTTCCGGAGCTCCGGAGCCACTACAGATCGTGTCAATTACTGACTATATC      539

Qy       309 CCAGACCCGCTGCCCCCGTTGCGGGGGGGCTTTCGCGGGAAGGCTTCGGC      368
             |||||||                 || ||  ||||| |     |||||
Db       540 TGAGATAAAGATCCTTTAATAACCAAGGCAGCAAAGAGGGCCTTCTCTCCAGTGAAGGGGA      599

Qy       369 CCAGGATAAGACCCGACTGGAAGCCTTTATCGAGGCCCAGCCGCTT-----CTGGCTCGAA      423
             |||                        | ||  ||| |    ||      ||| ||||
Db       600 ACTG--AAAGACCAGCTCGAGAACTTTCGCCGGGATCCAAATATTTCGAGTTGGCTGGAG      657

Qy       424 GACTATGCGCTCTTTATGTGCCCTCAAGCACCCGTTTGACGGGCAAGCCCTGGAACGAGTGG      483
             |   ||  ||  |||  |  ||||     ||||  || | ||| ||  ||| || |||||
Db       658 GATGCTGCTTATTTTGCTGCCATAGACAACTCTGTAAACACTATTAGCTGGTATGATTGG      717

Qy       484 AGCCCCGGAGCTGCGCGACCGTGAACCGGTGCCCTGCCCGCAGGGCGCCGTGAGGAGCTGGGCC      543
             |||    || || ||||||||  || | || || |||   ||| |||   ||||||||
Db       718 CCTGAACCATTGAAAAAATCGCCATCTGCAGCTCTAGAAGAGTTTATCAAAGTGAAAAG      777

Qy       544 GAGGAGGTGGCCCTTTACGAGTGGATTCAGTGGCGTTTTTTATCTGGAATGGGCCAGACC      603
             |   ||  ||          ||  || |||         |  || | ||  |||||
Db       778 GATTTATTGACATAATCATTGCACACAGTTGTTCTTGTTCCAACGACAAGTGAAAAAGTT      837

Qy       604 AAGGCCTATGCCGAATCCAAGGGGATTCAGATTATCGCGATATGCCCATCTTTGTTGGCC      663
             |||   |||||| |||||||    | || |   |  ||| ||||  | ||| ||||||
Db       838 CGTGACTATGCAGATCGCACGATCCAAAGGAATCAGTATATATGTTGGCCAATATATGTTGGA      897

Qy       664 TTCGATTCCTCAGATGTCTGAGGGCCAACCCGCAGTACTTCTACCCGAGGGCCATGGCAAC      723
             |||  || |     |||||||||| |     || |||| |  | ||||  |||| |||
Db       898 TATCACAGTGCTGATGTTTGGGCCCAACAAGAAACAATTTTGCTGAATAGGAAAAGGTTTC      957

Qy       724 CCCACGGTGGTGCGGGGTTCCGGCGGTTCCGGACTACTTCTCCGGGGACTACTTCAACCTATGGGGC      783
             |||                  ||   || |  |   |  |||    ||   || ||
Db       958 CCTCTTATAGTTAGTGGTGGTGTTGGTGGTGGTGTTGGTGGTTCAACTGTGAAACTGTGAAAAC      1017

Qy       784 AATCCGCTCTATCGCTCGGATGTGATGAAACGGACAACTTTGCCTGGTGCATTGCCCGC      843
             |||  ||| | ||  || |||| ||| |   |  ||| ||||||  || ||
Db      1018 AGCCCTCTTCTATGATTGGAAAGCCATGGAAGGATGAATTTTCATGTGGGTACGCCGA      1077

Qy       844 ATAAGGCAGTGCCTCGTCAAGCAGTGCCACCTGGTGCGCATCGACGCACTTTCGCGCGGGTTTGAA      903
             ||              ||     || |  |  |||| || ||||| ||||||| | ||
Db      1078 ATTCAACGTGCAACGGATCTTTTTGATGAATATTAGGATAGATCACTTTAGAGGATTTGCT      1137
```

```
Qy       904 GCCTACTGGGAGGTTCCGTTTGCCGCGGCCCCAATGCTGTGGAGGGCGGCTGGGTCAAAGCC       963
             |||||||||||||   ||||||||  || ||||  |||| |||||      ||  ||||
Db      1138 GGATTTGGGCGTTCCTTCCTTCTGAGGAAAAAAATTGCAATTCTCTGGAGACGGTGGAAGGTGGGA      1197

Qy       964 CCAGGGGAGAAGCTGTTGCTGCGCGGGTGCGGGCCCAACTGAGCGATGCGCCCATCATTGCC      1023
             |||||||||    |||||||||||   | |||  ||  | |||| |||| |  |||
Db      1198 CCTGGAAAGCCTTTGTTGATGCATCTATCTTACAGAGCTGTGGAGGATCAATATTATAGCA      1257

Qy      1024 GAAGACCTGGGGGTGATCACCCCCGAGGTGGAGCTTTGCGCGATGGTCGGGTTCCCC      1083
             ||||              |||  |||     |||   |     |||  |   || |
Db      1258 GAAGACTTGGGAGTAATTACCGGAGGACGTTGTTCAGCTAAGAAGTCCATTGAGGCACCT      1317

Qy      1084 GGCATGAAGATTTGCAGTTTGCTTTCTTTCGGTGAGGACAACGCCCTTTTGCCCCACAAC      1143
             |||     |    || ||||    |     ||| | || | ||||||  || |  ||  | |
Db      1318 GGAATGGCTGTGTACTCCAGTTTGCATTTGGCAGTGACGCTGAAAACCCTCATTACCTCAC      1377

Qy      1144 TACCCGCGCACGGCAATGTCGTGGTGTAACGGCGGAACCCACGACACGACACCACCCTG      1203
             ||          || | |   |||  || |        ||| || || ||||||
Db      1378 AATCATGAGCA--GAACCAAGTAGTGTATACTGGACACATGACACATGACAGATCCCGA      1434

Qy      1204 GGATGGTTCCGCACCGGCGCCGGAGGCCGAGCGGGGCCTTCATGCGGCCTACCTGGCCCGC      1263
             ||      || ||    |  |  ||      ||  || || || |  ||  || |
Db      1435 GGTTGGTGGGATACTTTGCCACAGAGAGAAATCCAATGTACTAAAGTATTATC----      1490

Qy      1264 TATGGCATCCGTGTTGTTGTCGGAATACAGAGGTCGCGGGCGCTTTGAGCTGGCCTTC      1323
             |||               || ||   |  ||  | |   || |||  |||| ||
Db      1491 ---------AAATATTGAGGAAGAGGAAATATCACGGGCTTGATCGAAGGTGCAGTT      1539

Qy      1324 AAAAGCCCGGCCAAGCTGGCTATTGTGCCTTTGCAGGACGTGCTGGGGCCCCCAG      1383
             ||||||  |     ||   || |  | || || ||  ||| ||| ||    ||||
Db      1540 TCTTCTGTAGCCCGTATTGCAATTATACCGATGCAGGATGTGCAGGATGTTCTTGGGGCTGAT      1599

Qy      1384 GCCCGCATGAACTTCCCCGACAGCGGCTGGGGGACCAACTGGGCGGTTGGCCGGCG      1430
             |||||  ||||  |  | |   || ||  | || || ||||| |
Db      1600 TCCAGAATGACAATTCCAGCAACATTTGGAAAACTGGAGCTGGGGAGTTGGAG      1646
```

RESULT 14
ABN69262
ID   ABN69262   standard; DNA; 1491 BP.
XX
AC   ABN69262;
XX
DT   01-JUL-2002 (first entry)
XX
DE   Streptococcus polynucleotide SEQ ID NO 6437.
XX
KW   Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW   group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW   antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS   Streptococcus pyogenes.
XX
PN   WO200234771-A2.
XX
PD   02-MAY-2002.
XX
PF   29-OCT-2001; 2001WO-GB004789.
XX
PR   27-OCT-2000; 2000GB-00026333.
PR   24-NOV-2000; 2000GB-0028727.
PR   07-MAR-2001; 2001GB-00005640.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Telford J, Masignani V, Margarit Y Ros I, Grandi G, Fraser C;
PI   Tettelin H;
XX
DR   WPI; 2002-352536/38.
DR   P-PSDB; ABP28631.
XX
PT   New Streptococcus protein for the treatment or prevention of infection or

PT   disease caused by Streptococcus bacteria, such as meningitis, and for
PT   detecting a compound that binds to the protein.
XX
PS   Claim 7; Page 3805; 4525pp; English.
XX
CC   The invention relates to a protein (ABP25413-ABP30895) from group B
CC   streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC   (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
CC   the specification. The proteins have antibacterial and antiinflammatory
CC   activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC   antibodies that bind (I) are used in the manufacture of medicaments for
CC   the treatment or prevention of infection or disease caused by
CC   Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC   Nucleic acids encoding (I) are used to detect Streptococcus in a
CC   biological sample. (I) is used to determine whether a compound binds to
CC   (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC   used as a vaccine or diagnostic composition. The disease caused by
CC   Streptococcus that is prevented or treated may be meningitis. Nucleic
CC   acid encoding (I) may be used to recombinantly produce (I) and may be
CC   used in gene therapy. Antibodies to (I) are used for affinity
CC   chromatography, immunoassays, and distinguishing/identifying
CC   Streptococcus proteins
XX
SQ   Sequence 1491 BP; 474 A; 247 C; 309 G; 461 T; 0 U; 0 Other;

Query Match   11.5%; Score 173; DB 6; Length 1491;
Best Local Similarity   47.1%;
Matches 567; Conservative   0; Mismatches 635; Indels   3; Gaps   1;

Qy   9 CCAACCGCGTTTGGAATTTGCTCCACCCCACCAGTTTCCGGGTCGCTGGGGGATTGG 68
           ||| ||||| || |||| || |||| ||||||||||||| | ||||||||||||
Db   6 CAAACGTGCAAGCGGTATCTTAATGCATATTAGTTCCCGCTGGGAAATTTGGATAGG 65

Qy   69 GGCTCTGGGCCCCGAGCGGTTTTTGGACTGCTGGCGGCGGGAGCCCGCTG 128
Db   66 CACTTTTGGAAAATCAGTCTTTTGAATTTGTTGATTTGATTTTGAATTGTTCTTAGCAGACGAGAAACAAACCTA 125

Qy   129 GTGGCAGGTCTTACCGCTGGGCCCTACCAGTTACGGCGACTCGCCGTACCAGTCCTTCTC 188
Db   126 TTGGCAAATTCGCCCTTTAACACGACGAGTTTTGGAGATTCCTTATCAGTCATTTTC 185

Qy   189 GGCTTTTGCGGTAACCCGTATTTGGTTGACCCCGAGATGCTGATTGAAAAAAGGCTGGCT 248
Db   186 AGCTATTGCTGGGAACACACATTTCATTGATTTTGAATTGCTAGTAGATGATGAACTTTT 245

Qy   249 GGAACAAAGCGAAGCGCCCCCGCGTATCCGACCCAGCGCGTGGATTATGGCTGGCTTTA 308
Db   246 AGAAGCAGCAGATTTGTGTGATATTACATTCCGGCACACAAATCCTGAAGCAGTAGACTATGC 305

Qy   309 CCAGACCCCGCTGGCCCCCTGTTGCGCGGCGGGCTTTCGCGGGGTTTCGGCGCAAGGGCTTCGGC 368
Db   306 TCAGCTTTTCAAGTTAGACGTCCACTTTTAGAAAAAGCAGTGAGAGCTTTTGTTGTCTGCTGA 365

Qy   369 CCAGGATAAGCAACCGACTGGAGAAGCCTTTATCGAGGCGGCCTTCGGCTTGGCTGGAAGACTA 428
Db   366 ACAAGAAAAATGTGTGTAAATTAGAAGCTTTTGAAACAGCTTCTAGCTGGTTAACTGATTT 425

Qy   429 TGCGCTCTTTATGGCCCTCAAGACCCCGGTTTGACGGCAAGCCCTGGAACAGAGTGGAGCCC 488
Db   426 TGCTGAATTTATGGCTTTAAAAGAATATTTTAACAATAAAGCCTTACAAGATTTGGGACGA 485

Qy   489 CGAGCTGCGC---GACCGTGAACCGGCTGCCCTGCCCCAGGGCCCGTGAGGAGCTCGGCCGA 545
Db   486 CGAAACTGTCATTAAACGCCAAGAAGACTCACTTAATAATTACCGTAACCGTGCTTGCAAA 545

Qy   546 GGAGGTGGCCCCTTTACGAGTGGATTCAGTCAGTGGCTTTTTATCTGGAATGGGCCACCAA 605
Db   546 AAAGATTACCTATCATAAAGTCTGTCAATACTTCTTTTACCAACAATGGTCAGCCCTGAA 605

Qy   606 GGCCATGCCGAATCCAAGGGAGATTCAGATTATCGGCGACTGATATGCCCATCTTTGTCGCCTT 665
Db   606 AACTTACGCTACAACCATAAAGGGATAGAAATTATTGGTGATAGACAAATGCCCTATTTATGTTTTCTGC 665

Qy   666 CGATTCCTTCAGATGTCTGGCCAACCCGCAGTACTTCTACCTCGAGGCCGATGGCAACCC 725
Db   666 GGACAGTGTAGAAGTTTGGACGATGCCTGCAGCTATTTAAAGTAGTAGATAGTGATAAAAAGCC 725

Qy   726 CACGGTGGTGGCGCGGCGTTCCGCGGGACTACTTCTCCGAAACCGGCCAGCTCTGGGCAA 785
Db   726 ATTGTTTATTGCAGGTGTTACCAGCAGATGGCTTTTAGTGAAGATGGTCAGCTGTGGGAAA 785

Qy   786 TCCGCCTCATCGCTGGGATGTGATGGAAAGGACAACTTGCCTGGTCATTGCCCGCAT 845
Db   786 TCCAACCTACAATTGGTCTGCCATGAAAAGTCTAATTTGCATGGTGGATTATCGTAT 845

Qy   846 AAGGCAGTGCTCAAGCAGCAGTGCCACCTGGTCGCCATCGACCACTTCCGCGGGTTTGAAGC 905
Db   846 TCAAGAAAGTTTCAAACTATATGATCAGCTAAGAATCGATCATTTTAAAGGCTTCTCAGA 905

Qy   906 CTACTGGGAGGTTCCGTTTGGCCGGCCCCAATGCTGTGGAGGGGCGCTGGGTCAAAGCCCC 965
Db   906 TTTTTGGGAAATTCCTGCTGGTGATAAAACTGCTAGAAATGGGCCATTGGCCTTCAGCCGC 965

Qy   966 AGGGGAGAAGCTGTTTGCTGCGGCCAACTGAGCGATGCGCCCATCATTGCCGA 1025
Db   966 AGTTATAGCCATTATTTTCTGCGGTTCGTGAGGCGGCGATGGTTTGCGCGATGGCTTCCCGG 1025

Qy   1026 AGACCTGGGGTGATCACCCCGAGGTGGAGGTTTGCGCGATGGCTTCCCCCG 1085
Db   1026 AAACCTGGGTTTATATTGATGAAAAGCAGAGCAGCTATTAGCATCAACAGGCTTTTCCTGG 1085

Qy   1086 CATGAAGATTTGCAGTTTGCTTTCCGGTGAGGACAACGCCTTTTGCCCCACAACTA 1145
Db   1086 CATGAAGATATTAGAATTTGGTTTATTTGATATAACGAGTCAAAGTATAGATTTACCACA 1145

Qy   1146 CCCCGCCACGGCAAATGGGTCGTGTAGCAGGAGCCCACCACCCCTGGG 1205
Db   1146 TTACTATGACCGTAATTGTCGTCGCTTACACAGGGACTCATGATAATAACGAGGGTGGTTAACGG 1205

Qy   1206 ATGGT 1210
          |||||
Db   1206 ATGGT 1210

RESULT 15
ABN69261
ID   ABN69261 standard; DNA; 1494 BP.
XX
AC   ABN69261;
XX
DT   01-JUL-2002 (first entry)
XX
DE   Streptococcus polynucleotide SEQ ID NO 6435.
XX
KW   Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW   group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW   antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS   Streptococcus agalactiae.
XX
PN   WO200234771-A2.
XX
PD   02-MAY-2002.
XX
PF   29-OCT-2001; 2001WO-GB004789.
XX
PR   27-OCT-2000; 2000GB-00026333.
PR   24-NOV-2000; 2000GB-00028727.
PR   07-MAR-2001; 2001GB-00005640.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;
PI   Tettelin H;
XX

```
DR   WPI; 2002-352536/38.
DR   P-PSDB; ABP28630.
XX
PT   New Streptococcus protein for the treatment or prevention of infection or
PT   disease caused by Streptococcus bacteria, such as meningitis, and for
PT   detecting a compound that binds to the protein.
XX
PS   Claim 7; Page 3805; 4525pp; English.
XX
CC   The invention relates to a protein (ABP25413-ABP30895) from group B
CC   streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC   (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
CC   the specification. The proteins have antibacterial and antiinflammatory
CC   activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC   antibodies that bind (I) are used in the manufacture of medicaments for
CC   the treatment or prevention of infection or disease caused by
CC   Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC   Nucleic acids encoding (I) are used to detect Streptococcus in a
CC   biological sample. (I) is used to determine whether a compound binds to
CC   (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC   used as a vaccine or diagnostic composition. The disease caused by
CC   Streptococcus that is prevented or treated may be meningitis. Nucleic
CC   acid encoding (I) may be used to recombinantly produce (I) and may be
CC   used in gene therapy. Antibodies to (I) are used for affinity
CC   chromatography, immunoassays, and distinguishing/identifying
CC   Streptococcus proteins
XX
SQ   Sequence 1494 BP; 494 A; 238 C; 316 G; 446 T; 0 U; 0 Other;

Query Match           11.3%;    Score 169.4;   DB 6;    Length 1494;
Best Local Similarity 48.9%;    Pred. No.:2.3e-30;
Matches  543;    Conservative   0;    Mismatches 561;    Indels  7;    Gaps  3;

Qy     11 AACGGCGCTTTGGAATTTGCTCCACCCACCAGTTTTCCGGGTCGCTGGGGGGGATTGGGG 70
Db      8 AACGTGCAAGTGGTGTCTTAATGCACATCACTTTCTCTACCAGTGATTTAGGTATTGGAA 67

Qy     71 CTCTGGGCGCCGAGGCCCAGCGCGGTTTTTGGACTGGCTGCCGATGCGGGAGCCCGCTGGT 130
Db     68 CATTTGGACGAGAAGCCTACGCTTTTGGTTGATTTTTCGTCGAAACTGATCAAAAATTT 127

Qy    131 GGCAGGTCTTACCGGTGGGCCCTACCAGTTACGGCGACTCGCCGTACCAGTCCTTCTCGG 190
Db    128 GGCAAATTCTTCCCTTAACAACAAGCTTCGGGGACTCTCCTTACCAGTCTTTTTCTG 187

Qy    191 CTTTTGCCCGTAACCCGTATTTGGTTGACCCCGAGATGCTGATTGAAAAAGGCTGGCTGG 250
Db    188 CAGTAGCTGGTAATACACATTTAATTGAATTTGATTTGCTAACGCTTGAAGGTTTCATTT 247

Qy    251 AACAAAGCGAAGCGGCCCCCGTATCCGCCGTATCCGCCAGCGCGTGGATTTGGCTGG 310
Db    248 CGAAAGATGA--CTACCAAAAATATTAGTTTTGGACAGGATCCAGAGAAGTTGTTGATTATGC 305

Qy    311 AGACCCGCTGGCCCCCTGTTGCGGCGGCGTTTCGCGGGGCGTTTCGGCGG 370
Db    306 TGGTTCTGTTTGAAAAACGCGTCCAGTTTTGGACAGGCTTCGGGCAAGGGCTTAAAAATTCTTGCAAGA 365

Qy    371 AGGATAAG--ACCCGACTGGAAGCCTTTATCGAGGCCGAGCGCTTCTGGCTGGAAGACTA 428
Db    366 AGAGAGAGCTACGAGAATGCTATCTGATTTCTTGCAAGAGAAGAAAATGGGTAACTGATTT 425

Qy    429 TGCGCTCTTTATGCCTTCAAGACCCGGTTGACGGCAAGCCCTGAACAGAGTGGAGCCC 488
Db    426 TGCTGAATTTATGGGGATCAAAGAACATTTGGTAATAACGGCGCTTCAAGAAATGGGATGA 485

Qy    489 CGAG--CTGCGCGACCGTGAACCGGCTGCCCTGGCCCAGGGCCGTGAGGAGCTGGCCGA 545
Db    486 CAAGGCTATTATACGCCGCGAAGAGAAGAAGCCTTAGCAGGATATCGTCAAAAGCTTAGTGA 545

Qy    546 GGAGGTGGCCCTTTACGAGTGGATTCAGTGGCTTTTTATCTGGAATGGGCCCAGACCAA 605
Db    546 AGTGGATAAAATATCATGGAAGTAACGCAATATTTCTTTACAAACAATGGTTTGAGGTTAAA 605

Qy    606 GGCCTATGCCGAATCCAAGGGGATTCAGATTATCGGGCGATATGCCCATCTTTGTGGCCTT 665
Db    606 AGAATATGCTAATGATAAAGGGATTCAAATTATCGGTGATATGCCCAATCTACGTTTCTGC 665

Qy    666 CGATTCCTCAGATGCTGTGGCCAACCCGCAGTACTTCTACCTCGAGGCCGATGGCAACCC 725
Db    666 CGATAGTGTAGAAGTTTGGACAATGCCTGAACTGTTTAAATTAGATCGGACAAACAACC 725

Qy    726 CACGGTCGTGTGGCGGGGCGTTCCGGCGGGACTACTTCTCCGAAACCGGCAGCTCTGGGGCA 785
Db    726 ACTTGCTATTGCTGTGGTACCAGCGGATGAGATTTTTCTGACGATGGTCAGCTTTGGGGCCAA 785

Qy    786 TCCGCTCTATCGCTGGGATGTGATGGAAAGGGACAACTTGCCTGGTGCATTGCCCGCAT 845
Db    786 TCCTATTTACAATTGGGATTACCACAAAGAGTCAGATTTTGACTGGTGGATTTATCGTAT 845

Qy    846 AAGGCAGTCGCTCAAGCAGTGCCACTGGTGCGGCATCGACCACTTCCCGGGGGTTGAAGC 905
Db    846 TCAATCTGGTGTTAAGATGTATGACTATTTGCCTATTGATCATTTTAAAGGTTTTTCAGA 905

Qy    906 CTACTGGGAGGTTCCGTTTGGCCGGCCCCAATGCTGTGGAGGGCGCTGGTCAAAGCCCC 965
Db    906 TTATTGGGAAATTCGAGAGATTATCAAACAGCTAATGATGGCTCATCGCAACCAGCTCC 965

Qy    966 AGGGGAGAAGCTGTTTGCTGCGGTCGGGCCCCACTGAGCGATGCGCCCATCATTGCCGA 1025
Db    966 CGGTCCAGAATTATTTGCAACTATCAAAGAAAAATAGGTGATTTACCTATTATTGCTGA 1025

Qy   1026 AGACCTGGGGGTGATCACCCCCGAGGTGGAGCGTTGTGCGCGATGGCTTCGGGTTCCCCGG 1085
Db   1026 GAATTTAGGCTATATTGATGAGAGAGAGCAGAGAGATTACTGGCTGGAACAGGGTTCCCAGG 1085

Qy   1086 CATGAAGATTTGCAGTTTGCTTTTTCCGGT 1116
Db   1086 TATGAAAATCATGGAATTCGGTTTTATGAT 1116
```

Search completed: January 14, 2006, 11:07:47
Job time : 888 secs

```
              GenCore version 5.1.6
        Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   January 10, 2006, 14:04:33 ; Search time 39 Seconds
                                        (without alignments)
                                       1236.015 Million cell updates/sec

Title:          US-10-618-976-2
Perfect score:  2714
Sequence:       1 MQLQRAFGILLHPTSFPGRW.........DLEPGLAAGLRALAEASQRA 501

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:  283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters:  283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  PIR_80:*
            1:  pir1:*
            2:  pir2:*
            3:  pir3:*
            4:  pir4:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1305.5	48.1	502	2	AH2289	4-alpha-glucanotra
2	1222.5	45.0	505	2	S74648	4-alpha-glucanotra
3	1194	44.0	505	2	E95246	4-alpha-glucanotra
4	1194	44.0	505	2	B98111	4-alpha-glucanotra
5	1100	40.5	485	2	E70363	4-alpha-glucanotra
6	1035.5	38.2	576	1	A45049	4-alpha-glucanotra
7	727	26.8	506	2	F70120	4-alpha-glucanotra
8	451	16.6	526	2	G72091	4-alpha-glucanotra
9	451	16.6	526	2	F86531	glucanotransferase
10	430.5	15.9	527	2	G71557	probable glucanotr
11	400.5	14.8	527	2	A81712	4-alpha-glucanotra
12	341.5	12.6	698	2	AC0016	4-alpha-glucanotra
13	320.5	11.8	724	2	G70928	probable 4-ALPHA-G
14	313.5	11.6	684	2	F83375	hypothetical prote
15	302.5	11.1	693	2	AI0996	4-alpha-glucanotra
16	302	11.1	694	1	C65137	4-alpha-glucanotra
17	300.5	11.1	738	2	T00748	4-alpha-glucanotra
18	293	10.8	694	2	A86007	4-alpha-glucanotra
19	293	10.8	694	2	B91161	4-alpha-glucanotra
20	292.5	10.8	726	2	B82511	4-alpha-glucanotra
21	271.5	10.0	699	2	H64118	4-alpha-glucanotra
22	249	9.2	489	2	H86711	hypothetical prote
23	122	4.5	1271	2	E83350	hypothetical membr
24	116	4.3	554	2	H95922	probable oxidoredu
25	114	4.2	585	2	T36060	probable oxidoredu
26	111	4.1	595	2	H95006	beta-galactosidase
27	110	4.1	595	2	C97879	beta-galactosidase
28	105.5	3.9	618	2	D87651	prolyl oligopeptid
29	104.5	3.9	629	2	E82624	thiamin biosynthes
30	103.5	3.8	600	2	H75516	probable maltoolig
31	103.5	3.8	751	2	T46517	hypothetical prote
32	102.5	3.7	996	2	AG2366	hypothetical prote
33	101.5	3.7	5232	2	A45086	HC-toxin synthetas
34	100	3.7	658	2	E95111	endo-beta-N-acetyl
35	99	3.6	721	2	C97980	endo-beta-N-acetyl
36	99	3.6	1361	2	T29435	hypothetical prote
37	98.5	3.6	599	2	D83764	chitinase BH0916 (
38	98.5	3.6	755	2	C90439	cation transportin
39	98.5	3.6	1072	2	A84112	alkaline amylopull
40	98	3.6	274	2	S04610	gene 2 protein - p
41	97.5	3.6	678	2	A65193	hypothetical 77.2
42	97.5	3.6	699	2	D70667	hypothetical prote
43	97.5	3.6	770	2	C87316	alpha-N-acetylgluc
44	97.5	3.6	1041	2	C83548	hypothetical prote
45	97	3.6	421	2	T43790	probable glutamate

ALIGNMENTS

```
RESULT 1
AH2289
4-alpha-glucanotransferase [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. (strain PCC 7120)
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120 #text_change 09-Jul-2004
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AH2289
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguch
  Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Yamada, M.; Tabata,
  DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ai
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AH2289
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-502 <KUR>
A;Cross-references: UNIPROT:Q8YQG0; UNIPARC:UPI0000CE92F; GB:BA000019; PIDN:BAB75570.:
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr3871
C;Superfamily: 4-alpha-glucanotransferase

Query Match         48.1%;  Score 1305.5;  DB 2;  Length 502;
Best Local Similarity  49.9%;  Pred. No. 5.1e-96;
Matches  252;  Conservative   80;  Mismatches 162;  Indels  11;  Gaps   5;

Qy     1 MQLQRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQ 60
           ||| ||  :|||||||| | |  |||    :| |: |:  |   :  |||||||| ||.|
Db     1 MPFIRSSGVLLHPTSCPSRFGIGDLGLEAYKFIDFLEKSYQQYWQVLPLGPTGYGNSPYM 60

Qy    61 SFSAFAGNPYIVDPEMLIEKGWLEQSEAP--PPYPTQRVDYGWLYQTRWPLLRRAFAGFR 118
         : ||:||| :::: |:|::: |     :    :| ::: ||    ::: :|| : :| :
Db    61 SYSALAGNHLLISPEKLLDEGLLSDEDFAHLPNFPNEKVDFDQVAPIKIQLLKKACENFK 120

Qy   119 ARASAQDKTRLEAFIEABRFWLEDYALFMALKTRFPDGKPWNEWSPELRDREPAALARARE 178
           ||  |:  ||  |  : ||||||||||||||| | ||         :|  ||: : :
Db   121 TKASPLQQKGFAGFCETKSYWLDDYALFMALKDTKDSSSWHTWEPALAKGREPDALEKVQR 180

Qy   179 ELABEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFL-E 237
           :  : :| :    || |:||| |  : :|| :|||||:|::| |:|:||||| |:| :
Db   181 QLTDEIFYYKFIQYEFFRQWSELKSYANMRGIEIIGDIPIYVAHDSADVWANPDIFCLDE 240

Qy   238 ADGNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHF 297
           : |  :: ||||:|||| |||||||| | | :  :|||| | ||:: :  ::||||||||
Db   241 ETGEVALMAGVPPDYFSATGQLWGNPVYNWEELQKQDFKWWVQRFEAMLDYVDVIRIDHF 300

Qy   298 RGFEAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDG 357
         ||||| | ||  |  : :| ||:|||| ||: ::::| .  ::|:|||||||||||||||:
Db   301 RGFEAFWTVPQGEETAMNGEWVTAPGEELFDAIKQKLGKLPVLAEDLGVITPEVEALRDK 360

Qy   358 FGFPGMKILQFAF-SGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMR 416
         : ||| | : |:  ||||||||||: || |||||||||||||||||||| | ||: :
```

```
Db    361  YEFPGMKVLQFAFGSDPGNPFLPNYSR--NFVVYTGTHDNDTTVGWNTANDYEKNNLL  418
Qy    417  AVLARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAW  476
Db    419  LYIG-----CISPEGIHWDLIRLALSSVANQAIVPLQDILGLGNEARMNFPSIAEGNWAW  473
Qy    477  RYAEGDLEPGLAAGLRALAEASQRA  501
Db    474  RYDFAVLTDELSDRLKILTKLYGRA  498

RESULT 2
S74648
4-alpha-glucanotransferase malQ - Synechocystis sp. (strain PCC 6803)
N;Alternate names: hypothetical protein sll1676
A;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C;Accession: S74648
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S74648
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-505 <KAN>
A;Cross-references: UNIPROT:P72785; UNIPARC:UPI000012EB5E; EMBL:D90900; GB:AB001339; NID
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Gene: malQ
C;Superfamily: 4-alpha-glucanotransferase

Query Match        45.0%;   Score 1222.5;  DB 2;   Length 505;
Best Local Similarity  48.6%;   Pred. No. 2.1e-89;
Matches  242;  Conservative  76;  Mismatches  169;  Indels  11;  Gaps  6;

Qy     4  QRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQSFS   63
Db     4  KRCSGILLHPTSLPSRFGIGDLDGAFQFIDFLADADQSVWQILPLGPTFGFGNSPYLCYS   63
Qy    64  AFAGNPYLVDPEMLIEKGWLEQS--EAPPYPTQRVDYGWLYQTRWPLLRRAFAGFRARA   121
Db    64  ALAINPWLIJSLDRLAEEGFLPPSLLDQAPPFTNPRVDYDQAIAYKSQVLKQAFAQFRTNI  123
Qy   122  SAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAREELA   181
Db   124  ELAIEQEFAEFCQAQSDWLADYALFMAIKEAHNGAGWHOWDKDIAWREPEALKIWGDRLK   183
Qy   182  EEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDWANPQYFYLEAD-G   240
Db   184  TEVLYHQFLQFLGFRQWQEVKAYANQRHIAIFGDLPIYVAHDSADVWANPENPCLDPETG   243
Qy   241  NPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGF  300
Db   244  EAAMMAGVPPDYFSATGQLWGNPVVDWETLKATGPAWWIKRFKANLQYLDIVIRDHFRGF  303
Qy   301  EAYWEVPFGRPNAVEGRWVKAPGEKLPAAVRAQLSD-APIIAEDLGVITPEVEALRDGFG  359
Db   304  ESYWGVPQGEKTAENGEWYPAPGKEFFQALGKALGDNLPIVAEDLGVITPEVEALRDEFN  363
Qy   360  FPGMKILQFAF-SGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAPMRAY  418
Db   364  FPGMKVLHFAFDSDRGNPLPFNY-SNGNAVVYTGTHDNDTTVGWFQERSEDDQQKVINY   422
Qy   419  LARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWRY  478
Db   423  LG-----CVCNEGIHWSLIRLASSVAALAIFPLQDILGLGSDCRMNLPGTAAGNWGWRY   477
Qy   479  AEGDLEPGLAAGLRALAE  496


Db   478  HPDQLNDWLSGHLSFITE  495

RESULT 3
E95246
4-alpha-glucanotransferase [imported] - Streptococcus pneumoniae (strain TIGR4)
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
C;Accession: E95246
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; He
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple.
nson, T.; Hickey, E.K.; Holt, I.B.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morriso
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: E95246
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-505 <KUR>
A;Cross-references: UNIPROT:P29851; UNIPARC:UPI0000051B0A; GB:AE005672; PIDN:AAK76166..
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP2107
C;Superfamily: 4-alpha-glucanotransferase

Query Match        44.0%;   Score 1194;   DB 2;   Length 505;
Best Local Similarity  48.1%;   Pred. No. 3.8e-87;
Matches  242;  Conservative  77;  Mismatches  168;  Indels  16;  Gaps  9;

Qy     4  QRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQSFS   63
Db     3  KRQSGVLMHISSLLPGAYGIGSFGQSAYDFVDFLVRTKQRYWQIILPLGATSYGDSPYQSFS  62
Qy    64  AFAGNPYLVDPEMLIEKGWLEQSEAPP---PYPTQRVDYGWLYQTRWPLLRRAFAGFRAR  120
Db    63  AFAGNTHFIDLDILVEQGLLEASDLEGVDFGSDASEVDYAKIYYARRPLLEKAVKRFFEV  122
Qy   121  ASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEW-SPELRDREPAALARAREE  179
Db   123  GDVKD---FEKFAQDNQSWLELFAEYMAIKEYFDNLAWTEWPDADARARAKASALESYREQ  179
Qy   180  LAEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDWANPQYFYLEAD  239
Db   180  LADKLVLYHRVTQYFFQQWLKLKAYANDNHIEIVGDMPIYVAEDSSDMWANPHLFKTDVN  239
Qy   240  GNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRG  299
Db   240  GKATCIAGCPPDEFSVTGQLWGNPIYDWEAMDKDGYKWWIERLRESFKIYDIVRIDHFRG  299
Qy   300  FEAYWEVPFGRPNAVEGRWVKAPGEKLPAAVRAQLSDAPIIAEDLGVITPEVEALRDGFG  359
Db   300  FESYWEIPAGSDTAAPGEWVKGPGYKLFAAVKEELGELNIIAEDLGFMTDEVIELRERTG  359
Qy   360  FPGMKILQAFSGEDNAF-LPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAY   418
Db   360  FPGMKILQFAFNPEDESIDSPHLAPA--NSVMYTGTHDNNTVLGWYRN--EIDDA-TREY  414
Qy   419  LARYGIRCLSEYE-VAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWR  477
Db   415  MARYTNR--KEYETVVHAMLRTVFSSVSFMAIATMQDLLELDEAARMNFPSTLGGNWSWR  472
Qy   478  YAEGDLEPGLAAGLRALAEASQR  500
Db   473  MTEDOLTPAVEEGLLDLTTIYRR  495

RESULT 4
B98111
4-alpha-glucanotransferase (EC 2.4.1.25) [imported] - Streptococcus pneumoniae (strain
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C;Accession: B98111
```

R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
J.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: B98111
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-505 <KUR>
A;Cross-references: UNIPROT:P29851; UNIPARC:UPI00000051B0A; GB:AE007317; PIDN:AAL00719.1;
C;Genetics:
A;Gene: malM
C;Superfamily: 4-alpha-glucanotransferase
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match   44.0%; Score 1194; DB 2; Length 505;
Best Local Similarity 48.1%; Pred. No. 3.8e-87;
Matches 242; Conservative 77; Mismatches 168; Indels 16; Gaps 9;

```
Qy    4 QRAFGILLHPTSFPPGRWGIGALGREABRFLDWLADAGARWWQVLPLGPTSYGDSPYQSFS 63
          |||:|:||||| ||| :||| |:||  |:   |  | :||| |:|||||| ||||||||||||
Db    3 KRQSGVLMHISSLPGAYGIGSFGQSAYDFVDFLVRTKQRYWQILPLGATSYGDSPYQSFS 62

Qy   64 AFAGNPYLVDPEMLIEKGWLEQSEAPP---PYPTQRVDYGWLYQTRWPLLRRAFAGFRAR 120
          ||||| |:     |:|||  ||  :     ::|||| |||| | :|| |||||||| ||||
Db   63 AFAGNTHFIDLDILVEQGLLEASDLEGVDFGSDASEVDYAKIYYARRPLLEKAVKRPFEV 122

Qy  121 ASAQDKTRLEAPIEAERFWLEDYALFMALKTRPDGKPWNEW-SPELRDREPAALARAREE 179
          |:|::|||| |:|||:|| ||:|:|||||     || ||  ||||||:||:|| |: |
Db  123 GDVKD---FEKPAQDNQSWLELFAEYMAIKEYFDNLAWTEWPDADARARKASALESYREQ 179

Qy  180 LAEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEAD 239
          | ::| |:|||:|||:|||||||||:||||||||||||||||||||||:|| || ::
Db  180 LADKLVYHRVTQYFFQQWLKLKAYANDNHIEIVGDMPIYVAEDSSDMWANPHLFKTDVN 239

Qy  240 GNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRG 299
          |||| || |||||| ||:||||| |  | |:||:|| ||| :|| ::|  |||||||||
Db  240 GKATCIAGCPPDEFSVTGQLWGNPIYDWEAMDKGYKWWIERLRESFKIYDIVRIDHFRG 299

Qy  300 FEAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFG 359
          | |||| :||||| |:|   | | :||| | | |: :|| ||||:|  :|:||||:|
Db  300 FESYWEIPAGSDTAAPGEWVKGPGYKLFAAVKEELGELNIIAEDLGFMTDEVIELRERTG 359

Qy  360 FPGMKIQFAFSGEDNAP-LPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAY 418
          ||||||||||| |:|| |:| | :||||:|:|||||||||:|:|||||:|  | ::||
Db  360 FPGMKIQFAFNPEDESIDSPHLAPA--NSVMYTGTHDNNTVLGWYRN--EIDDA-TREY 414

Qy  419 LARYGIRCLSEYE-VAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWR 477
          ||  ||| ||:: |  | :|:||||| ||:  ||||  || |:||:||| ||||  |:|
Db  415 MARYTNR--KEYETVHAMLRTVFSSSVSFMAIATMQDLLELDEAARMNFPSTLGGNWSWR 472

Qy  478 YAEGDLEPGLAAGLRALAEASQR 500
          ||||||  |  ||  :|:::| |
Db  473 MTEDQLTPAVEEGLLDLTTIYRR 495
```

RESULT 5
E70363
4-alpha-glucanotransferase (amylomaltase) - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C;Accession: E70363
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: E70363
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-485 <AQF>

A;Cross-references: UNIPROT:O66937; UNIPARC:UPI0000056423; GB:AE000704; NID:g2983301;
A;Experimental source: strain VF5
C;Genetics:
A;Gene: malM
C;Superfamily: 4-alpha-glucanotransferase

Query Match   40.5%; Score 1100; DB 2; Length 485;
Best Local Similarity 45.8%; Pred. No. 1.1e-79;
Matches 219; Conservative 74; Mismatches 169; Indels 16; Gaps 6;

```
Qy    5 RAPGILLHPTSFPPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSY--GDSPYQSF 62
          | |||||| ||| :|||| ||::||| ||| |  ||  :|||||| ||  |   |:|| |
Db    2 RLAGILLHVTSLPSPYGIGDLGKEAYRFLDFLKECGFSLWQVLPNPTSLEAGNSPYSSN 61

Qy   63 SAFAGNPYLVDPEMLIEKGWLEQSEAPPYPTQRVDYGWLYQTRWPLLRRAFAGFRARAS 122
          | |||| ||:||| ||::||:|:|  :  |:   | :| |::: :||::|| |:||
Db   62 SLFAGNYVLIDPEELLEEDLIKERDL-KRFPLGEALYEVVYEYKKELLEKAFKNFRR--- 117

Qy  123 AQDKTRLEAFIEAERFWLEDYALFMALKTRPDGKPWNEWSPELRDREPAALARAREELAE 182
             |::||  |  |:|||:|||:|| |:||   | | |||  ||||:||:| |||||| |
Db  118 ---FELLEDLPLKEHSYWLRDYALYMAIKEE-EGKEWYEWDEELKRREKEALKRVLNKLKG 173

Qy  183 EVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADGNP 242
           :  ::|::|:|: : | ::|  :| || ||||:|:| :|:||||:||  :| ||| :|
Db  174 RFYFHVFVQFVFFKQWEKLRRYARERGISIVGDLPMYPSYSSADVWTNPELFKLDGDLKP 233

Qy  243 TVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGFEA 302
           :|||||  | ||GQLWGNP:  || :|:| || | || : | :||  | |||||||||
Db  234 LFVAGVPPDFFSKTGQLWGNPVTNWEEHEKEGFRWWIRRVHHNLKLFDFLRLDHFRGFEA 293

Qy  303 YWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGFPG 362
          |||||:| : |:.|||||||| :||  :  : ||| |||||||:||:||  || || ||
Db  294 YWEVPYGEETAVNGRWVKAPGKTLFKKLLSYFPKNPFIAEDLGFITDEVRYLRETFKIPG 353

Qy  363 MKILQFAFSGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAYLARY 422
           :::.||| |:|  LPHN  ||:|  | ::||||  | | ||| |||::| :| ||| |
Db  354 SRVIEFAFYDKESEHLPHN--VEENNVYTSTHDLPPIRGWFENLGEESRKRLFEYLGRE 411

Qy  423 GIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWRYAE 480
            :   |:|::|  | |:|:|| | ||:|:| ||:|| ||||:| ||| |::| |||:||
Db  412 ----IKEEKVNEELIRLVLISRAKFAIIQMQDLLNLGNEARMNYPGRPFGNWRWRIKE 465
```

RESULT 6
A45049
4-alpha-glucanotransferase (EC 2.4.1.25) - potato
N;Alternate names: amylomaltase; D-enzyme; disproportionating enzyme
C;Species: Solanum tuberosum (potato)
A;Note: sequence extracted from NCBI backbone (NCBIN:122077, NCBIP:122078)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: A45049
R;Takaha, T.; Yanase, M.; Okada, S.; Smith, S.M.
J. Biol. Chem. 268, 1391-1396, 1993
A;Title: Disproportionating enzyme (4-alpha-glucanotransferase; EC 2.4.1.25) of potato.
A;Reference number: A45049; MUID:93123262; PMID:7678257
A;Accession: A45049
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-576 <TAK>
A;Cross-references: UNIPROT:Q06801; UNIPARC:UPI0000129746; GB:X68664; GB:S52648; NID:g;
A;Experimental source: cv. May Queen, tubers
C;Superfamily: 4-alpha-glucanotransferase
C;Keywords: chloroplast; glycosyltransferase; hexosyltransferase

Query Match   38.2%; Score 1035.5; DB 1; Length 576;
Best Local Similarity 44.9%; Pred. No. 1.9e-74;
Matches 216; Conservative 72; Mismatches 178; Indels 15; Gaps 6;

```
Qy    4 QRAFGILLHPTSFPPGRWGIGALGREABRFLDWLADAGARWWQVLPLGPT----SYGDSPY 59
          :||:|||||||||| :|||| ||  || ||||||  || :||||||| |     ||:|||
Db   77 RRRAGILLHPTSFPGPYGIGDLGPQAFKFLDWLHLAGCSLWQVLPLVPPGKRGNEDGSPY 136

Qy   60 QSFSAFAGNPYLVDPEMLIEKGWLEQSEAPPYPTQRVDYGWLYQTRWPLLRRAFAGFRA 119
```

```
Db   137  SGQDANCGNTLLISLEELVDDGLLKMELPEPLPTDRVNYSTISEIKDPLITKA-AKRLL 195

Qy   120  RASAQDKTRLEAFIEAERF--WLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAR 177
                |||||: : : :   |    ||||||||:::|    | | :|  :
Db   196  SSEGELKDQLENFRRDPNISSWLEDAAYFAAIDNSVNTISWYDWPEPLKNRHLAALEEVY 255

Qy   178  EELAEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLE 237
               |   :::|| :   | :|||||||:   ||||||||:   |
Db   256  QSEKDFIDIFIAQQFLFQRQWKKVRDYARSKGISIMGDMPIYVGYHSADVWANKKQFLLN 315

Qy   238  ADGNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHF 297
               ||||||| |||:  |||:|||:|||:|:   |||||||:|||||||:|   :|:|
Db   316  RKGFPLIVSGVPPDAFSETGQLWGSPLYDWKAMEKDGFSWWVRIQRATDLFDEFRIDHF 375

Qy   298  RGFEAYWEVPFGRPNAVEGRWVTKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDG 357
               ||||:|||| | |||| ||||:||||||||||||||:::|||||||: |   ::
Db   376  RGFAGFWAVPSSEKIAILGRWKVGPGKPLFDAILQAVGKINIIAEDLGVITEDVVQLRKS 435

Qy   358  FGFPGMKILQFAF-SGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEABRAFMR 416
               ||||||||:   |  :|||  |||:|| |||:|||||||||:  |:||||:|:|:|
Db   436  IEAPGMAVLQFAFGSDAENPHLPHNH--EQNQVVYTGTHDNDTIRGWWDTLPQEEKSNVL 493

Qy   417  AYLARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAW 476
               |:||:|||| |  :|||||||| ||||:::|   :  :||
Db   494  KYLSN-----IEEEISRGLIEGAVSSVARIAIIPMQDVLGLGSDSRMNIPATQFGNWSW 548

Qy   477  R 477
          |
Db   549  R 549
```

RESULT 7
F70120
4-alpha-glucanotransferase (malQ) homolog - Lyme disease spirochete
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C;Accession: F70120
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb, J.F.; Flei
schmann, R.D.; Richardson, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S
.; Hanson, M.; van Vugt, R.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman, J.; Utterback,
T.; Watthey, L.; McDonald, L.; Artiach, P.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.
D.; Horst, K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943; PMID:9403685
A;Accession: F70120
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-506 <KLE>
A;Cross-references: UNIPROT:O51188; UNIPARC:UPI0000057391; GB:AE001127; GB:AE000783; NID
A;Experimental source: strain B31
C;Superfamily: 4-alpha-glucanotransferase

```
Query Match   26.8%;  Score 727;  DB 2;  Length 506;
Best Local Similarity  33.1%;  Pred. No. 5.6e-50;
Matches  173;  Conservative  87;  Mismatches  216;  Indels  46;  Gaps  8;

Qy    1  MQLQRAFGILLHPTSFPGRWIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDS-PY 59
              :|  :|||| ::||  | | :|    |||| |       :||| |||||||||:
Db   10  LNLKRKSGILLNISSLPSKYGIGDLGKGAYKFIDFLFASSQSYWQMFAYSPIDFTRSPPY 69

Qy   60  QSPSAFAGNPYLVDPEML----IEKGWLEQSEAPPYPTQRVDYGWLYQTRW------- 107
              ||||||||| |      | :    |:|| |
Db   70  SIFSAFAGNVYYIDLEALDKFIDSDLSLLKENE----------TRYSDLKKIS 112

Qy  108  ---PLLRRAFGFRARASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWS-- 162
                :||:|||:||| | :|||:||||||||||||||||||||||||
Db  113  FKDKFLKEAALNFINRASADEVRSFEKFKKKSSYWLLDFASFVAFKEYFFKESKNAFNVL 172

Qy  163  --PELRDREPAALARAREELAEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFV 220
                ||:|||||| ||||: ||||||||||||||| :| |||||:: :|
Db  173  FDRGILKRNEKDLFKLRNILSEIKVQEVLQYFFSQFQALKRYANDKGIELIMNVPFFI 232
```

RESULT 8
G72091
4-alpha-glucanotransferase CP0431 [imported] - Chlamydophila pneumoniae (strains CWL02
C;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004
C;Accession: G72091; F81577
R;Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.
; Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey
, E.; Peterson, J.; Utterback, T.; Berry, K.; Bass, S.; Linher, K.; Weidman, J.; Khoury
, H.; Small, D.; Catanese, J.; Coyle, K.; Jonasen, J.; Freeman, N.; Heidelberg, J.;
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; McClarty, G.; Salzber
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzber
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A;Reference number: A81500; MUID:20150255; PMID:10684935
A;Accession: F81577
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-526 <REA>
A;Cross-references: UNIPARC:UPI0000047BEF; GB:AE002204; GB:AE002161; NID:g71895350; PID
A;Experimental source: strain AR39, HL cells
C;Genetics:
A;Gene: malQ; CP0431
C;Superfamily: 4-alpha-glucanotransferase

```
Query Match   16.6%;  Score 451;  DB 2;  Length 526;
Best Local Similarity  26.8%;  Pred. No. 5.3e-28;
Matches  131;  Conservative  87;  Mismatches  212;  Indels  58;  Gaps  15;

Qy   21  GIGALGREAERFLD------------WLADAGARWWQVLPLGPTSYGDSPYQSFSAFAGNPYLVDP 74
              |||:|| :||||:             |||||| ||||||||||||| :||:|| |||| ||||||
Db   41  GIG--------EFLDLIPLISWCQKQGFSVIQLLPLNDTGEDTSPYNSISSVALNPLFLSL 93

Qy   75  EML---------IEKGWLEQSEAPPYPTQRVDYGWLYQTRW 107
                  |      :|| || ||  |||||| |:    |
Db   94  SSLPNIDTIPEVAKKLQDMHEL----CSTPSVSYTQVKEKKWAFLREYYQKC-CKSSLEGN 149

Qy  127  TRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAREELAEEVAL 186
              |||||||||| ||||||:||| |||||||:  |:   |||:|||| |:|:|||||||
Db  150  SNFSEFLSSERYWLYPYGTFRAIKHHMGEPINWPKSLTDQE--NFPDLTKKFHDEVLF 207

Qy  187  YEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADGNPTVVA 246
              :|| |||| |||| |||||:|| |||||||||||:||||  |||||:    |
Db  208  FSYIQFLCYQQLCEVKAYADQHHVLLKGDLPLILISKDSCDVWYFRDYF------SSSRSV 261

Qy  247  GVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGFEAYWEV 306
```
(column 2 / right side:)
```
Db  221  AFPDSSDVWANPQYFYLEADGNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIA 280
              |:::|:|   ||:::|         :|| ::||:||:|:::::::: : |
Db  233  AYDSADVWAYQKYFKLRFDASKDKIAGISPDYFLEQEQAWDSAAYSWNVLKKFKYEWWAK 292

Qy  281  RIRQSLKQCHLVRIDHFRGFEAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPII 340
              ::|::|:  |:||||||| |||||  | ::||| :::  |  :|| |
Db  293  RIGVLRKYADIIKIDHFRGFVSTWEVSAGESYAFNGLWVKSPGRDFFNFILNEIKDLKIW 352

Qy  341  AEDLGVITPEVEALRDGFGPGMKILQFAFS-GEDNAFLPHNYPAHGNVVVYSGTHDNDT 399
              |||||||||:|   | | |||||||| :     |||:|||:||||:::|   :| |
Db  353  VEDFQNDLEDVSRLRDFFNFPGMKIMNLAFDFDSSNQNLPHNYIK--NCIVITGSGDNDT 410

Qy  400  TLGWFRTAPEABRAFMRAYLARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLG 459
              |:|||::::: :::::: :: :|:|:
Db  411  IREFINSLDDLHKKYIFDYLN-----TNENFVWDMIRSAMGSVSDNVIVSMQDYINLG 464

Qy  460  PEARMNFPGRLGDNWAWRYAEGDLBPGLAAGLRALARASQRA 501
              |  :||:| |::|   :::::::: ::|::|:::|:|::|:|
Db  465  DKPSANIPKSTLDNCIFRLLESDLDATLSQNISFITRLYGRA 506
```

RESULT 9

F86531

glucanotransferase [imported] - Chlamydophila pneumoniae (strain J138)

C;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: F86531
R;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is
Nucleic Acids Res. 28, 2311-2314, 2000
A;Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A;Reference number: A86491; MUID:20030349; PMID:10871362
A;Accession: F86531
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-526 <STO>
A;Cross-references: UNIPROT:Q9Z8L2; UNIPARC:UPI0000047BEF; GB:BA000008; NID:g8978700; PI
C;Genetics:
A;Gene: malQ
C;Superfamily: 4-alpha-glucanotransferase

Query Match   16.6%; Score 451; DB 2; Length 526;
Best Local Similarity 26.8%; Pred. No. 5.3e-28;
Matches 131; Conservative 87; Mismatches 212; Indels 58; Gaps 15;

```
Qy    21 GIGALGREAERFLD------WLADAGARWWQVLPLGPTSYGDSPYQSFSAFAGNPYLVDP 74
            |||:| :  ||:            ||||:|| |||||:||: || : | |||| ||||:
Db    41 GIG-------EFLDLPLISWCQKQGFSVIQLLPLNDTGEDTSPYNSISSVALNPLFLSL 93

Qy    75 EML--------IEKGWLEQSEAPPYPTQRVDYGWLYQTRWPLLRRAFAGFRARASAQDK 126
                        ::|||:|||  | :  ::| |:  :||:  ||::||::  |:|
Db    94 SSLPNIDTIPEVAKKLQDMHEL---CSTPSVSYTQVKEKKWAFLREYYQKC-CKSSLEGN 149

Qy   127 TRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAREELAEEVAL 186
          :|| ||:|||:||  |  ::|   |:| |:||||  |||::| |:  |:||||||:|
Db   150 SNFSEFLESERYWLYPYGTFRAIKHMHGEPINNWPKSLTDQE-----NFPDLTKKFHDEVLF 207

Qy   187 YEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLBEADGNPTVA 246
          : |:|| | :| : ||||:| |||:||||||||||||||||||| | :| |   || ||
Db   208 FSYLQPLCYQQLCEVKAYADQHHVLLJKGDLPILISKDSCDVWYFRDYF-----SSSRSV 261

Qy   247 GVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGFEAYWEV 306
          ||| || || |:||:|| | : :::|| |||||||||||||||||||||: | | |:
Db   262 GAPPDLYNSEGQNWHLPIYNFSQLAKDDYIWWKERLRYAQNFYSVRLDHIIGFFRLWI- 320

Qy   307 PFGRPNAVEGRWVK------APGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFG 359
           :|| ||  ||: ||       || |:||:||||||||||||||||||||||||||||
Db   321 ---WDSSGRGRFIPDNPKDYIKQGTEILSTMLGASSMLP-IGEDLGIIPQDVKTLTHLG 376

Qy   360 FPGMKILQFAFSGE-DNAFLPHNYPAHGNVVYSGTHDNDTTLGWFRTAPEAERAFMRAY 418
            || :|  : ||| :: |||:||:|:  :|| | |||| | | || :|||   ||:||:
Db   377 ICGTRIPRWERNWESDSAFIPLKDYNPLSVTTLS-THDSDTFAQWMLNSPKEAKQFAKFL 435

Qy   419 LARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPE-------ARMNFPGRL 470
           |  | : |:| |:|  | : | |:   | |  | |:|::| ||        ||||||:|
Db   436 HLPFQKTLTTETQI--DILKLSHESASIFHINLFNDYLALCPDLVSKNLQRERINTPGTI 493

Qy   471 G-DNWAWR 477
          | ||:|:|
Db   494 SKKNWSYR 501
```

RESULT 10

G71557

probable glucanotransferase - Chlamydia trachomatis (serotype D, strain UW3/Cx)

C;Species: Chlamydia trachomatis
C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 09-Jul-2004
C;Accession: G71557
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell
Science 282, 754-759, 1998
A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tra
A;Reference number: A71570; MUID:99000809; PMID:9784136
A;Accession: G71557
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-527 <ARN>
A;Cross-references: UNIPROT:O84089; UNIPARC:UPI0000047C0C; GB:AE001283; GB:AE001273; N]
A;Experimental source: serotype D, strain Uw-3/Cx
C;Genetics:
A;Gene: malQ
C;Superfamily: 4-alpha-glucanotransferase

Query Match   15.9%; Score 430.5; DB 2; Length 527;
Best Local Similarity 25.9%; Pred. No. 2.3e-26;
Matches 126; Conservative 89; Mismatches 233; Indels 39; Gaps 12;

```
Qy    33 LDWLADAGARWWQVLPLGPTSYGDSPYQSFSAFAGNPYLVDPEMLIEKGWLEQSEAPPY 92
          :||  :||:  |  : :   |  || ||   | ||||:: |  ::|:||:|||||||:
Db    56 IDWCISCGFQZLQILPINDTGSCSSPYNSISSIALNPLHLSISALPYKEEVPAAETRIRE 115

Qy    93 PTQRVDYGWLYQTRWPLLRRAFAGFRARASAQ---DKTRLEAFIEAERFWLEDYALFMA 148
             :  :   :  :| |:: |  | ::  |:|    ||:  :|  : :: :||  ||:|
Db   116 MQQLSQLPQVHYEKVRSMKRDFFQEYYRVCKQKKLTDHPDFYAFCEQEKYWLHPYALFRS 175

Qy   149 LKTRFDGKPWNEWSPELRDREPAALARAREELAEEVALYEWIQWLFYLEWGQTKAYAESK 208
          :: : :  |:| |   : : |     : : ::  ::  ||:    |  : :|||:
Db   176 IREHLDNLPINHWPTTYTDL--SQITEHRTFAEDIQPHSYLQYLCFQQMTQVREHANCK 233

Qy   209 GIQIIGDMPIFVAFDSSDVWANPQYFYLBEADGNPTVVAGVPRDYFSETGQLWGNPLYRWD 268
          ||:| || || : :|| |||  | : |:      :: : | |||:| : ||:|  |  :|
Db   234 SCLIKGDIPILISKDSCDVWFYRHYF-----SSSESVGAPPDLYNAEGQNWHLPICNMK 287

Qy   269 VMERDNFAWCIARIRQSLKQCHLVRIDHFRGFEAYW---BVPFGRPNAVEGRWVKAPGEK 325
          :::|:| :| : ::|:  :   | || || | | |    :   | | :  | :  ::
Db   288 TLQQDNYLWWKERLRYAENFYSLYRLDHVVGLFRFWWDESGCGRFBHDPKNYLAQGQD 347

Qy   326 LFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGPGMKILQFAFSGEDN-AFLPHNYPA 384
          :: :: ||| :| | ||||| | |:|: :: | |  :| ::| :| |: ::  ::
Db   348 ILSHLLTSSSMLP-IGEDLGTIPSDVKRMLESFAVCGTRIPRWERNWGNGAYTP--FDQ 404

Qy   385 HGNVVYS-GTHDNDTTLGWFRTAPEAERAFMR----AYLARYGIRCLSEYEVAGALIEL 439
          : | ||: || : || | |: : |:|::     :|  |  | :  :: :| : ||:|
Db   405 YDPLSVTSLSTHDSSTLASWWKESPQESKLFAQFLGLPYSSTLSLHNHTE-----ILKL 458

Qy   440 AFKSPAKLAIVPLQDVLGLGPE---------ARMNFPGRLG-DNWAWRYABGDLEPGLAAG 490
          :: :  ::  : :  :|| ||          :| | |  :  :|:  :  | | |:::
Db   459 SHKTSSIFRINLINDYLALFPDLISKTPRYERINLPGTISKNNWVRVKPSIEDLSHSK 518

Qy   491 LRALAEA 497
          | :| ||
Db   519 LNSLLEA 525
```

RESULT 11

A81712

4-alpha-glucanotransferase TC0362 [imported] - Chlamydia muridarum (strain Nigg)

C;Species: Chlamydia muridarum, Chlamydia trachomatis MoPn

```
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C;Accession: A81712
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A;Reference number: A81500; MUID:20150255; PMID:10684935
A;Accession: A81712
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-527 <TET>
A;Cross-references: UNIPROT:Q9PKU9; UNIPARC:UPI000000578C9; GB:AE002303; GB:AE002160; NID
A;Experimental source: strain Nigg (MoPn)
C;Genetics:
A;Gene: TC0362
C;Superfamily: 4-alpha-glucanotransferase

Query Match          14.8%;   Score 400.5;  DB 2;   Length 527;
Best Local Similarity 25.9%;  Pred. No. 5.6e-24;
Matches 124;  Conservative  88;  Mismatches 205;  Indels  61;  Gaps  16;

Qy    33 LDWLADAGARWWQVLPLGPTSYGDSPYQSFSAFAGNPYLVDPEMLIEKGWLEQSEAPPPY  92
            ||| |    | || |             |   |   |
Db    56 IDWCTLCGFQILQILPINDTGSCSSPYNSISISLNP----------LHLSISALPV 102

Qy    93 --------PTQRVDYGWLYQTRW----PLLRRAFAGFRARASAQDKTR--LEAFI 133
                  :   : :              :: :
Db   103 KEEVSSSRKLIQEMQRLSQ--LSQVNYEKVIPMKRAFFKEYFRVCKSKNLTNHPDFCDFC 160

Qy   134 EAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARARERELAEEVALYEWIQWL 193
Db   161 EREKYWLHPYALFCSIREHLNYLPINHWSTTYTDL--SYISQHEHTFAKDIEFYSYLQYL 218

Qy   194 FYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADGNPTVVAGVPRDYF 253
Db   219 CFEQMKQVRKHADHKGCLIKGDIPILISKDSCDVWFYRKYF------SSSESVGSPPDFY 272

Qy   254 SETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGFEAYM---EVPFGR 310
Db   273 NAEGQNWNLPIYNMKTLRQDAYHWWKERLRYAENFYSLYRLDHVVGLFRFWVWDELGRGR 332

Qy   311 PNAVDEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGFPGMKILQFAF 370
Db   333 FEPQDPKDYLDQGTDILSHLLKASSMLP-IGEDLGTIPVDVKQALESLAVCGTRIPRWER 391

Qy   371 SGE-DNAFLPHNYPAHGNVVVYS-GTHDNDTTLGWFRTAPEAERAFMRAYLARYGIRCLS 428
Db   392 DWEGTGAYIP--FDQYNPLSVTSLSTHDSSTLALWWQEAPQEARLFAQ-FLGMPYTPSLS 448

Qy   429 EYEVAGALIELAFKSPAKLAIVPLQDVLGLGP-------EARMNFPGRLG-DNWAWR 477
Db   449 -FHNHKEILKLSHKTSSIFHINLINDYLALCPDLISTNPLQERINLPGTISKNNWVYR 505

RESULT 12
AC0016
4-alpha-glucanotransferase (EC 2.4.1.25) [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C;Accession: AC0016
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AC0016
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-698 <KUR>
A;Cross-references: UNIPROT:Q82JI1; UNIPARC:UPI00000DC99D; GB:AL590842; PIDN:CAC88989.1;
C;Genetics:
A;Gene: malQ
```

```
C;Superfamily: 4-alpha-glucanotransferase
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match          12.6%;   Score 341.5;  DB 2;   Length 698;
Best Local Similarity 26.0%;  Pred. No. 4e-19;
Matches 140;  Conservative  76;  Mismatches 245;  Indels  77;  Gaps  23;

Qy    20 WGIGALGREAERFLDWLADAGARWWQVLP---LGPTS-YGDSPYQSFSAFAGNPYLVDPE  75
            |||||                  |||
Db   162 WGIGDFG-DLKQMLEQVGERGSFIGLNPIHALYPANPHSASPYSPSSRRWLNVIYIDIN 220

Qy    76 MLIE-----KGWLEQSEAPPPYPTQR---VDYGWLYQTRWPLLRRAFAGFRARASAQ 124
                               |           |
Db   221 RVEEFQQSEAAQRWWHQAETQQALAKSRESEWVDYPLVMQLKLTALRLSFPLFTARKAKD 280

Qy   125 DKTR--LEAFIEAERFWLEDYALFMALKTRF-DGKP----WNEWSPELRDREPAALARARE 178
Db   281 AQVQAFRHFVEQGGSSLHQQAVFDALHAHLSEHDPMMWGMPVWPEKYRDGHSSAVADFCR 340

Qy   179 ELAEEVALYEWIQWLFYLEWGQTKAYAESK--GIQIIGDMPIFVAFDSSDVWANPQYFYL 236
Db   341 EHADEVTFYLWLQWLAASQPDDCFQSSQQRKMPIGLYRDLAVGVAEGGAETWCDRELYCL 400

Qy   237 EADGNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDH 296
Db   401 KAS------VGAPPDILGPLGQNWGLPPMDPHVMVARAYQPFIDLLRANMTSCGALRIDH 454

Qy   297 FRGFEAYWEVPFGRPNAVEGRWVKAPGEKLFA--AVRAQLSDAPIIAEDLGVITPE-VEA 353
Db   455 VMALLRLIWIPYGH-TADQGAYVKYPVDDLLAVLALESQRHHCMVIGEDLGTVPVEIVGK 513

Qy   354 LRDGFGFPPGMKILQFAFSGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRT------- 406
Db   514 LRDS-GVYSYKVLYFEHDSENIFRAPQSYPVQAMATI--TTHDLPTLRGYWQADDLITLGN 570

Qy   407 ----APEAERAFMRAYL---ARYG-IRCLSEYEV--AGALIELAFKSPA------KL 447
Db   571 KLGLYPD-QQILKQLYLDRERAKQGLLEGLHRYDCVPKKVGHKAALLSMSPVLNRGLQRY 629

Qy   448 AIVPLQDVLGLGPE-------ARMNFPGRLGDNWAWRYAEGDLEPGLAAGLRALAEASQ 499
Db   630 VADSASALLGLQPEDWLDMAAPVNIPGTTDEYPNWRRK----------LSASLEEIFADSQ 680

RESULT 13
G70928
probable 4-ALPHA-GLUCANOTRANSFERASE - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
A;Accession: G70928
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: G70928
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-724 <COL>
A;Cross-references: UNIPROT:O53932; UNIPARC:UPI000012EBSC; GB:AL022021; GB:AL123456; NI
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv1781c
C;Superfamily: 4-alpha-glucanotransferase

Query Match          11.8%;   Score 320.5;  DB 2;   Length 724;
Best Local Similarity 27.2%;  Pred. No. 2e-17;
Matches 153;  Conservative  69;  Mismatches 258;  Indels  83;  Gaps  26;

Qy     4 QRAFGILLHPTSFPGR-WGIGALGREAERFLDWLADA-GARWWQVLPL-----GPTSY  54
           :  :: :                    |||| ||||      |  |||       ||
```

```
Db   172  RRAWGLAVQLYSVRSRQSWGIGDLTDLANLAL-WSASAHGAGYVLVNPLHAATLPGPAGR   230
Qy    55  G----DSPYQSFSAFAGNP-YL--VDPEM--LIEKGWLEQSEAPPYPTQRVDYGWLYQ    104
Db   231  SKPIEPSPYLPTSRRFVNPLYLRVEAIPELVDLPKCRGVQRLRTNVQQHADQLD-TIDRD   289
Qy   105  TRWPLLRRAF-AGFRARASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSP   163
Db   290  SAWAAKRAALKLVHRVPRSAGRELAYAAFRTREGRALDDFATWCALAETY-GDDWHRWPK   348
Qy   164  ELRDREPAALARAREELAEEVALYEWIQWLF-YLEWGQTKAYAESKGIQIIGDMPIFVA   221
Db   349  SLRHPDASGVADFVDKHADAVDFHRWLQWQLDEQLASAQSQALRAGMSLGIMADLAVGVH   408
Qy   222  FDSSDVWANPQYFYLEADGNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIAR   281
Db   409  PNGADAWALQDVL--AQG---VTAGAPPDEFNQLGQDWSQPPWRPDRLAEQEYRPFRAL   462
Qy   282  IRQSLKQCHLVRIDHFRGFEAYWEVPFGRPNAVEGRWVKAPGEKL--FAAVRAQLSDAPI   339
Db   463  IQAALRHAGAVRIDHIIGLFRLMWIPDGAP--PTQGTYVRYDHDAMIGIVALEAHRAGAVV   521
Qy   340  IAEDLGVITPEVEALRDGF---GFPGMKILQFAFSGEDNAFLPHNYPAHGN-----VVVY   391
Db   522  VGEDLGTVEPWV---RDYLLLRGLLGTSILWFE---QDRDCGPAGTPLPAERWREYCLSS   575
Qy   392  SGTHDNDTTLGWFR---------------TAP--------EAERAFMRAYLARYGIRC---   426
Db   576  VTTHDLPPTAGYLAGDQVRLRESLGLLITNPVEAELESARADRAAWMAELRRVGLLADGAE   635
Qy   427  LSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGD--NWAWRYAEGDLE   484
Db   636  PDSEEAVLALYRYLGRTPSRLLAVALTD--AVGDRRTQNQPGTTDEYPNWRVPLTGPDGQ   693

Qy   485  PGLAAGL------RALAEASQRA   501
Db   694  PMLLEDIFTDRRAATLAEAVRAA   716
```

RESULT 14
F83375
hypothetical protein PA2163 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C;Accession: F83375
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: F83375
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-684 <STO>
A;Cross-references: UNIPROT:Q9I1V2; UNIPARC:UPI00000C5E0; GB:AE004643; GB:AE004091; NID
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA2163
C;Superfamily: 4-alpha-glucanotransferase

```
Query Match             11.6%;  Score 313.5;  DB 2;  Length 684;
Best Local Similarity   25.2%;  Pred. No. 6.6e-17;
Matches  145;  Conservative  70;  Mismatches 258;  Indels 103;  Gaps  23;

Qy     4  QRAFGILLHPTSFPG--------RWGI----------GALGREA--ERPLDWLADA    39
Db   125  QREWAVAMAPPSCPSLASLAPGRSRRWGLAAQVYALRRPGDGGLGDSAALEDLLRSAARH   184
Qy    40  GARWWQVLPL----GPTSYGDSPYQSFSAFAGNPYLVDPEMLIEKGWLEQ-------SE    87
Db   185  GADALAISPLHALAEANGHAYSPYSPSSRLFFNVLHAAPATILGAAAVEQAIRRAGLAAE   244
Qy    88  APPPYPTQRVDYGWLYQTRWPLLRRAFAGFRARASAQDKTRLEAFIEAERFWLEDYALFM   147
Db   245  MARLESLELIDWTAAADLRMRLLRQLHRDFTERASPLRHDLAEFRREAGEALLH-HCRFE   303
Qy   148  ALKTRFDGKP-WNEW-SPELRDREPAALARAREELAAEVALYEWIQWLFY--LEWGQTKA   203
Db   304  TLQAHLGAGPDWRRWPEPLRRPGEPAVAAFCADH-AEEVDFRAFGQWLTQRCLQHAQRQA   362
Qy   204  YAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADGNPTVVAGVPRDYFPSETGQLWGNP  263
Db   363  REAGMAIGLVADLAVGADGGGSQAWSRQEELLAE-----VNVGAPPDILNQSGQDWGVS   416
Qy   264  LYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGFEAYWEVPFGRPNAVEGRWVKAPG   323
Db   417  AFNPEGLRRHGYRAFREMLRANLAWPGGLRIDHVMGLQRLWLIPRGQPHA-GAYLRYPQ   475
Qy   324  EKL--FAAVRAQLSDAPIIAEDLGVITPEVEALRDGFG--PPGMKILQFAFSGEDNAFL   378
Db   476  RELLRLLALEASRASALVIGEDLGTVP--EGLREELARRQVLGTRVLLFERRGE--RFV   530
Qy   379  -PHNYPAHGNVVYSGTHDNDTTLGWFRTAPERAERAFMRAYLARYGIR--CLSEYEVA   433
Db   531  PPAQWPA--DAMATTSTHDLPSLSGWNRGRDIHWR------GRAGHRSAEECAADLELR   581
Qy   434  G-----------------------------ALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPG   468
Db   582  AEERRALAASLEPPVDPDPAAEVPLDACIGYVGATPAPLVLLPLEDALGSLEQPNLPGPG   641
Qy   469  RLGDNAWRYAEGDLE----PGLAAGLRALARASOR   500
Db   642  DAHPNWRRRWPENAAQMLGTPQVDRRLR-LLERSRR   676
```

RESULT 15
AI0996
4-alpha-glucanotransferase (EC 2.4.1.25) [imported] - Salmonella enterica subsp. enter...
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AI0996
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churche
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farra
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica ser
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AI0996
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-693 <PAR>
A;Cross-references: UNIPARC:UPI000005A7AC; GB:AL513382; PIDN:CAD08101.1; PID:g16505080.
C;Genetics:
A;Gene: malQ
C;Superfamily: 4-alpha-glucanotransferase
C;Keywords: glycosyltransferase; hexosyltransferase

```
Query Match             11.1%;  Score 302.5;  DB 2;  Length 693;
Best Local Similarity   24.7%;  Pred. No. 5e-16;
Matches  135;  Conservative  68;  Mismatches 238;  Indels 105;  Gaps  22;

Qy    20  WGIGALGREAERFLDWLADAGA------RWWQVLPLGPTSYGDSPYQSFSAFAGNPYLVDP   74
Db   157  WGIGDFGDLKSMLVDVATRGGAFPIGLNPIHALYPANPBS--ASPYSPSSRRWLNVIYIDV   214
Qy    75  EMLIEKGWLEQSEAPPYP----------TQRVDYGWLYQTRWPLLRRAFAGFRARASA   123
Db   215  NAVEDFRLSERAQAWWQMPATQQKLRQARDAQWVDYATVTALKITALRMAWTRFAARDDA   274
Qy   124  QDKTRLEAFIEAB---RFWLEDYALFMALKTRFDGKPWN--EWSPELRDREPAALARARE   178
Db   275  Q-MABFRHFIAREGESLYWQAAPDALHAYQVKEDGQRMGWPAWPEAYQVESPAVKQFCE   333
```

```
Qy   179  ELAEEVALYEWIQWLFYLE----WGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYF  234
               ||| ||||     ||||||     :::    |: ::  ||     :::  |  |
Db   334  AHREEVEFYLWLQWLAWRQFAACWDTCQSF--KLPIGLYRDLAVGVAEGGAETWCDRELY  391

Qy   235  YLEADGNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRI  294
               |:|        |||||  ::    :::   | :::|  |   :|| |:|:  | |
Db   392  CLKAS------VGAPPDILGPLGQNWGLPPMDPHIIVARAYEPFIDLLRANMQNCGALRI  445

Qy   295  DHFRGFEAYWEVPFGRPNAVEGRWVKAPGEKLFA--AVRAQLSDAPIIAEDLGVITPE-V  351
          ||  : | :  :  :|  :  :    :::: :   | :||:  :     |||| :|: |
Db   446  DHVMSLLRLWWIPYDE-TADQGAYVHYPVDDLLSILALESQRHRCMVIGEDLGTVPVEIV  504

Qy   352  EALRDGFGFPGMKILQFAFSGEDNAFLPHNYPAHGNVVVYSGTHD---------NDTTL  401
           |||| |             :|                |||||   |:   |||
Db   505  GKLRDS-GVYSYKVLWFENDLEKNFRAPGAYPQQSMAV--ASTHDLPTLRGYWECGDLTL  561

Qy   402  ----------GWFRTAPEAERAFMRAYLARYGIRCLSEYEVAGALIELAFKSPAKL  447
                      |  :: | | :  ||  :|| |||     :||    :|
Db   562  GKALGLYPDEVILRGLYEDRERAKQGLLDA-LHKYG--CLPKRAGHKALL--------M  609

Qy   448  --------QDVLGLGPE------ARMNFPGRLGDNWAWRYAEGDLEPGLA  488
                   ||:|| |         ||||       ::  ||     |:
Db   610  SMTPTLNRGLQRYIADSNSGLLGLQPEDWLDMADPVNVPGTSDQYKNWRRK------LS  662

Qy   489  AGLRAL  494
          | || |:
Db   663  ASLEAM  668
```

Search completed: January 10, 2006, 14:13:28
Job time : 41 secs

```
         GenCore version 5.1.6
         Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    January 10, 2006, 14:03:37 ; Search time 164 Seconds
                                        (without alignments)
                                        2155.303 Million cell updates/sec

Title:         US-10-618-976-2
Perfect score: 2714
Sequence:      1 MQLQRAFGILLHPTSFPGRW.........DLEPGLAAGLRALAEASQRA 501

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters:      2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      UniProt_05.80:*
                1: uniprot_sprot:*
                2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1696.5	62.5	500	1	MALQ_THETH	O87172 thermus the
2	1692.5	62.4	500	2	Q5SIV3_THET8	Q5siv3 thermus the
3	1692.5	62.4	500	2	Q72J82_THET2	Q72j82 thermus the
4	1635.5	60.3	500	2	Q6JHX9_THEAQ	Q6jhx9 thermus aqu
5	1626.5	59.9	500	2	Q6JHY0_THESC	Q6jhy0 thermus sco
6	1475	54.3	503	2	Q4HAI6_9DEIO	Q4hai6 deinococcus
7	1318	48.6	499	2	Q5N4Q2_SYNP6	Q5n4q2 synechococc
8	1305.5	48.1	502	2	Q8YQG0_ANASP	Q8yqg0 anabaena sp
9	1248	46.0	504	2	Q7NG57_GLOVI	Q7ng57 gloeobacter
10	1245	45.9	493	2	Q74DY3_GEOSL	Q74dy3 geobacter s
11	1222.5	45.0	505	1	MALQ_SYNY3	P72785 synechocyst
12	1221	45.0	518	2	Q4NUB1_9DELT	Q4nub1 anaeromyxob
13	1217	44.8	518	2	Q8DKZ1_SYNEL	Q8dkz1 synechococc
14	1194	44.0	505	1	MALQ_STRPN	P0a3q0 streptococc
15	1194	44.0	505	1	MALQ_STRR6	P0a3q1 streptococc
16	1174	43.3	497	2	Q8XHY6_CLOPE	Q8xhy6 clostridium
17	1154	42.5	505	2	Q7UT23_RHOBA	Q7ut23 rhodopirell
18	1117.5	41.2	494	2	Q608L7_METCA	Q608l7 methylococc
19	1108.5	40.8	487	1	MALQ_CLOBU	Q59266 clostridium
20	1108.5	40.8	540	2	Q73JN0_TREDE	Q73jn0 treponema d
21	1100	40.5	485	1	MALQ_AQUAE	O66937 aquifex aeo
22	1056	38.9	497	2	Q8P0K6_STRP8	Q8p0k6 streptococc
23	1055	38.9	497	2	Q5XBP8_STRP6	Q5xbp8 streptococc
24	1041	38.4	497	2	Q99ZC2_STRPY	Q99zc2 streptococc
25	1039	38.3	497	2	Q8K751_STRP3	Q8k751 streptococc
26	1035.5	38.2	576	1	DPEP_SOLTU	Q06801 solanum tub
27	1024.5	37.7	503	2	Q726G0_DESVH	Q726g0 desulfovibr
28	1023.5	37.7	499	2	Q5LZT4_STRT1	Q5lzt4 streptococc
29	1022.5	37.7	502	2	Q5M4E4_STRT2	Q5m4e4 streptococc
30	1020	37.6	498	2	Q8DYN6_STRA5	Q8dyn6 streptococc
31	1020	37.6	498	2	Q8E494_STRA3	Q8e494 streptococc
32	1016	37.4	601	2	Q6PYY7_OSTTA	Q6pyy7 ostreococcu
33	1009	37.2	556	2	Q5V0X8_HALMA	Q5v0x8 haloarcula
34	1007.5	37.1	576	2	Q9LV91_ARATH	Q9lv91 arabidopsis
35	1006.5	37.1	468	2	Q8ZXM0_PYRAE	Q8zxm0 pyrobaculum
36	1001.5	36.9	493	2	Q7P7Q5_FUSNV	Q7p7q5 fusobacteri
37	986	36.3	585	2	Q9FDV9_CHLRE	Q9fdv9 chlamydomon
38	982.5	36.2	509	2	Q8DT30_STRMU	Q8dt30 streptococc
39	976.5	36.0	506	2	Q8RF60_FUSNN	Q8rf60 fusobacteri
40	923.5	34.0	586	2	Q8LI30_ORYSA	Q8li30 oryza sativ
41	875.5	32.3	487	2	Q5NHN0_FRATT	Q5nhn0 francisella
42	841	31.0	516	2	Q7VBH6_PROMA	Q7vbh6 prochloroco
43	839.5	30.9	512	2	Q7UTL9_SYNPX	Q7utl9 synechococc
44	825	30.4	521	2	Q7V6S5_PROMM	Q7v6s5 prochloroco
45	796.5	29.3	506	2	Q7V1I0_PROMP	Q7v1i0 prochloroco

ALIGNMENTS

```
RESULT 1
MALQ_THETH
ID   MALQ_THETH              STANDARD;      PRT;   500 AA.
AC   O87172;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   4-alpha-glucanotransferase (EC 2.4.1.25) (Amylomaltase)
DE   (Disproportionating enzyme) (D-enzyme).
GN   Name=malQ;
OS   Thermus thermophilus.
OC   Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OC   Thermus.
OX   NCBI_TaxID=274;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=ATCC 33923 / DSM 674 / AT-62;
RA   Terada Y., Fujii K., Takaha T., Okada S.;
RT   "Cloning, expression and characterization of amylomaltase from Thermus
RT   aquaticus ATCC33923.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS)
RX   MEDLINE=20143895; PubMed=10677288; DOI=10.1006/jmbi.1999.3503;
RA   Przylas I., Tomoo K., Terada Y., Takaha T., Fujii K., Saenger W.,
RA   Strater N.;
RT   "Crystal structure of amylomaltase from Thermus aquaticus, a
RT   glycosyltransferase catalysing the production of large cyclic
RT   glucans.";
RL   J. Mol. Biol. 296:873-886(2000).
CC   -!- CATALYTIC ACTIVITY: Transfers a segment of a 1,4-alpha-D-glucan to
CC       a new position in an acceptor, which may be glucose or a 1,4-
CC       alpha-D-glucan.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the disproportionating enzyme family.
CC   --------------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   --------------------------------------------------------------------------
DR   EMBL; AB016244; BAA33728.1; -; Genomic_DNA.
DR   PDB; 1CWY; X-ray; A=1-500.
DR   PDB; 1ESW; X-ray; A=1-500.
DR   PDB; 1FP8; X-ray; A=1-500.
DR   PDB; 1FP9; X-ray; A=1-500.
DR   InterPro; IPR003385; Glyco_hydro_77.
DR   Pfam; PF02446; Glyco_hydro_77; 1.
DR   TIGRFAMs; TIGR00217; malQ; 1.
DR   3D-structure; Carbohydrate metabolism; Glycosyltransferase;
KW   Transferase.
FT   STRAND        6     10
FT   HELIX        13     15
```

```
Query Match             62.5%;  Score 1696.5;  DB 1;  Length 500;
Best Local Similarity   62.9%;  Pred. No. 5.2e-120;
Matches  315;  Conservative   65;  Mismatches  118;  Indels   3;  Gaps   2;

Qy    1 MQLQRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWQVLPLGPTSYGDSPYQ 60
          :|:||||||:||||||:||| :|:||:|||||::||: :|:||::|||||||||||||||
Db    1 MELPRAFGLLLHPTSLPGPYGVGVLGREARDFLRFLKEAGGRYWQVLPLGPTSYGDSPYQ 60

Qy   61 SFSAFAGNPYLVDPEMLIEKGWLEQSEAPPPYPTQRVDYGWLYQTRWPLLRRAFAGFRAR 120
          |||||||||||:| ::|:|:|::||:||    :||:|||||:|||:|||:|| | |:|:|:
Db   61 SFSAFAGNPYLIDLRPLAERGYVRLED--PGFPQGRVDYGLLYAWKMPALKEAFRGFKEK 118

Qy  121 ASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAREEL 180
          ||: :: ||||:| |||::||||||||||||   |||||:|||:|:|||||:||:|| |:
Db  119 ASPEERRAFAAFREREAWWLEDYALFMALKGAHGGLPWNRWPLPLRKREEKALREAKSAL 178

Qy  181 AEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADG 240
          |||||::|: ||||:|: || |: ::: :|||||||||||||:|||:||:|| ||:|:||
Db  179 AEEVAFHAFTQWLFFRQWGALKABABALGIRIIGDMPIFVAEDSAEVWAHPEWFHLDEEG 238

Qy  241 NPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGF 300
          :|||||||| |||||||||:|||||||||::|::||:  :|  ::| || ||||||||||
Db  239 RPTVVAGVPPDYFSETGQRWGNPLYRWDVLEREGFSFWIRRLEKALELFHLVRIDHFRGF 298

Qy  301 EAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGF 360
          |||||:|      ||||||||||||||  : :  |  ||::||||||||||||||||:||:
Db  299 EAYWEIPASCFTAVEGRWVKAPGEKLFQKIQEVFGEVPVLAEDLGVITPEVEALRDRFGL 358

Qy  361 PGMKILQFAF-SGEDNAFLPHNYPAHGNVVYSGTHDNDTTLGWFRTAPEAERAFMRAYL 419
          ||||:||||| :||:| ||||||||||:||: ||||||||||||:|||| |:||||:||
Db  359 PGMKVLQFAFDDGMENPFLPHNYPAHGRVVVYTGTHDNDTTLGWYRTATPHEKAFMARYL 418

Qy  420 ARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWRYA 479
          | :||  |||: | | : |||::|| ||| ||:|||| || |||||:|||  ||||||  :
Db  419 ADWGITFREEEEVPWALMHLGMKSVARLAVYPVQDVLALGSSARMNYPGRPSGNWAWRLL 478

Qy  480 EGDLEPGLAAGLRALAEASQR 500
           |:|  ||| |:||:|||| :|
Db  479 PGELSPEHGARLRAMAEATER 499
```

RESULT 2
Q5SIV3_THET8

```
ID   Q5SIV3_THET8  PRELIMINARY;      PRT;    500 AA.
AC   Q5SIV3;
DT   01-FEB-2005 (TrEMBLrel. 29, Created)
DT   01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT   01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE   4-alpha-glucanotransferase (Amylomaltase) (Disproportionating enzyme)
DE   (D-enzyme).
GN   OrderedLocusNames=TTHA1261;
OS   Thermus thermophilus (strain HB8 / ATCC 27634 / DSM 579).
OC   Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OC   Thermus.
OX   NCBI_TaxID=300852;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=HB8;
RA   Masui R., Kurokawa K., Nakagawa N., Tokunaga F., Koyama Y.,
RA   Shibata T., Oshima T., Yokoyama S., Yasunaga T., Kuramitsu S.;
RT   "Complete genome sequence of Thermus thermophilus HB8.";
RL   Submitted (NOV-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP008226; BAD71084.1; -; Genomic_DNA.
DR   SMR; Q5SIV3; 1-500.
DR   GO; GO:0004134; F:4-alpha-glucanotransferase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR   InterPro; IPR003385; Glyco_hydro_77.
DR   Pfam; PF02446; Glyco_hydro_77; 1.
DR   TIGRFAMs; TIGR00217; malQ; 1.
KW   Complete proteome; Transferase.
SQ   SEQUENCE   500 AA;  57194 MW;  C4B5A3EDD9DC1336 CRC64;
```

```
FT   HELIX          27     38
FT   TURN           39     40
FT   STRAND         43     45
FT   STRAND         52     52
FT   TURN           54     55
FT   TURN           58     59
FT   STRAND         61     61
FT   HELIX          69     71
FT   HELIX          75     79
FT   TURN           80     81
FT   STRAND         94     94
FT   HELIX          97    118
FT   HELIX         121    150
FT   TURN          151    153
FT   HELIX         156    158
FT   HELIX         161    164
FT   HELIX         165    166
FT   TURN          168    177
FT   TURN          178    178
FT   HELIX         179    205
FT   TURN          206    207
FT   STRAND        209    215
FT   HELIX         223    227
FT   HELIX         229    231
FT   HELIX         232    232
FT   STRAND        234    234
FT   TURN          236    237
FT   STRAND        240    240
FT   STRAND        242    243
FT   STRAND        245    246
FT   STRAND        256    257
FT   STRAND        263    263
FT   HELIX         265    270
FT   TURN          271    273
FT   STRAND        274    286
FT   HELIX         289    293
FT   STRAND        294    294
FT   HELIX         295    298
FT   TURN          299    299
FT   STRAND        300    305
FT   TURN          306    307
FT   TURN          312    313
FT   STRAND        315    318
FT   HELIX         321    332
FT   STRAND        337    339
FT   TURN          342    343
FT   HELIX         347    355
FT   TURN          356    357
FT   STRAND        360    363
FT   HELIX         364    366
FT   TURN          372    373
FT   TURN          375    376
FT   HELIX         378    380
FT   HELIX         383    384
FT   STRAND        387    390
FT   TURN          394    395
FT   TURN          399    405
FT   HELIX         408    420
FT   TURN          421    422
FT   HELIX         428    430
FT   HELIX         431    441
FT   STRAND        446    450
FT   HELIX         451    454
FT   HELIX         455    456
FT   TURN          459    461
FT   HELIX         466    467
FT   TURN          472    473
FT   STRAND        475    475
FT   TURN          479    480
FT   HELIX         484    496
FT   TURN          497    498
SQ   SEQUENCE    500 AA;  57222 MW;  6E7175A38E0A5D61 CRC64;
```

Query Match
Best Local Similarity 62.4%; Score 1692.5; DB 2; Length 500;
                       62.7%; Pred. No. 1e-119;
Matches 314; Conservative   66; Mismatches 118; Indels 3; Gaps 2;

```
Qy    1 MQLQRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQ   60
          : |:||||||||||::|: |:|: :||:|||||||| ||| ::||||||||||||||
Db    1 MELPRAFGLLLHPTSLPGPYGVGVLGQEARDFLRFLKEAGGRYWQVLPLGPTGYGDSPYQ   60

Qy   61 SFSAFAGNPYLVDPEMLIEKGWLEQSEAPPYPTQRVDYGWLYQTRWPLLRRAFAGFRAR  120
         |||||||||||| ||||:|: :|:  :  |:  ||  :::|::||||||||  |||
Db   61 SFSAFAGNPYLIDLRPLAERGYVRLED--PGFPQGRVDYGLLYAWKWPALKEAFRGFKEK  118

Qy  121 ASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAREEL  180
         |: ::: || ||: |:|:||:: : :|:|||||:   ||::|| : : :  |: :
Db  119 ASPEREAFAAFRREREAWWLEDYALFMALKGAHGGLPWNRWPLPLRKREEKALREAKSAL  178

Qy  181 AEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADG  240
         |||||: ||||:||||||||| ||: ::|| |||||||||| ||:: |||| :||:|||
Db  179 AEEVAFHAFTQWLFFRQWGALKAEAEALGIRIIGDMPIFVAEDSAEVWAHPEWFHLDEEG  238

Qy  241 NPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGF  300
         ||||||||| ||||||| |||||||||| ||: |:  :: :|:|:| : ||||||||||
Db  239 RPTVVAGVPPDYFSETGQRWGNPLYRWDVLEREGFSWIRRLEKALELFHLVRIDHFRGF   298

Qy  301 EAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGF  360
         |||| ||||| |||||||||||||||:  :: |: ::|| |||||||||||||:|| ||
Db  299 EAYWEIPASCPTAVEGRWVKAPGEKLFQKIQEVFGEVPVLAEDLGVITPEVEALRDRFGL  358

Qy  361 PGMKILQFAF-SGEDNAFLPHNYPAHGNVVVVSGTHDNDTTLGWFRTAPEAERAFMRAYL  419
         ||||:|||| ||::| |||||||||||:  :|||||||||||||||:||| ||||||
Db  359 PGMKVLQFAFDDGMENPFLPHNYPAHGRVVVVTGTHDNDTTLGWYRTATPHEKAFMARYL  418

Qy  420 ARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWRYA  479
         |:|||::||||||| ||| | | ||| ||: ||:|| |||:|:|||||  |||:||
Db  419 ADWGITFREEEVPWALMHLGMKSVARLAVYPVQDVLALGSEARMNYPGRPSGNWAWRLL   478

Qy  480 EGDLEPGLAAGLRALAEASQR  500
         ||||||||:||| |||:|
Db  479 PGELSPEHGARLRAMAEATER  499
```

RESULT 3
Q72J82_THET2
ID   Q72J82_THET2   PRELIMINARY;   PRT;   500 AA.
AC   Q72J82;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   4-alpha-glucanotransferase (EC 2.4.1.25).
GN   OrderedLocusNames=TTC0897;
OS   Thermus thermophilus (strain HB27 / ATCC BAA-163 / DSM 7039).
OC   Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OC   Thermus.
OX   NCBI_TaxID=262724;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   PubMed=15064768; DOI=10.1038/nbt956;
RA   Henne A., Brueggemann H., Raasch C., Wiezer A., Hartsch T.,
RA   Liesegang H., Johann A., Lienard T., Gohl O., Martinez-Arias R.,
RA   Jacobi C., Starkuviene V., Schlenczeck S., Dencker S., Huber R.,
RA   Klenk H.-P., Kramer W., Merkl R., Gottschalk G., Fritz H.-J.;
RT   "The genome sequence of the extreme thermophile Thermus
RT   thermophilus.";
RL   Nat. Biotechnol. 22:547-553(2004).
     EMBL; AE017304; AAS81241.1; -; Genomic_DNA.
     SMR; Q72J82; 1-500.
     GO; GO:0004134; F:4-alpha-glucanotransferase activity; IEA.
     GO; GO:0005975; P:carbohydrate metabolism; IEA.
     InterPro; IPR003385; Glyco_hydro_77.
DR   Pfam; PF02446; Glyco_hydro_77; 1.
DR   TIGRFAMs; TIGR00217; malQ; 1.
KW   Complete proteome.
SQ   SEQUENCE   500 AA;   57194 MW; C4B5A3EDD9DC1336 CRC64;

Query Match
Best Local Similarity 62.4%; Score 1692.5; DB 2; Length 500;
                       62.7%; Pred. No. 1e-119;
Matches 314; Conservative   66; Mismatches 118; Indels 3; Gaps 2;

```
Qy    1 MQLQRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQ   60
          : |:||||||||||::|: |:|: :||:|||||||| ||| ::||||||||||||||
Db    1 MELPRAFGLLLHPTSLPGPYGVGVLGQEARDFLRFLKEAGGRYWQVLPLGPTGYGDSPYQ   60

Qy   61 SFSAFAGNPYLVDPEMLIEKGWLEQSEAPPYPTQRVDYGWLYQTRWPLLRRAFAGFRAR  120
         |||||||||||| ||||:|: :|:  :  |:  ||  :::|::||||||||  |||
Db   61 SFSAFAGNPYLIDLRPLAERGYVRLED--PGFPQGRVDYGLLYAWKWPALKEAFRGFKEK  118

Qy  121 ASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAREEL  180
         |: ::: || ||: |:|:||:: : :|:|||||:   ||::|| : : :  |: :
Db  119 ASPEREAFAAFRREREAWWLEDYALFMALKGAHGGLPWNRWPLPLRKREEKALREAKSAL  178

Qy  181 AEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADG  240
         |||||: ||||:||||||||| ||: ::|| |||||||||| ||:: |||| :||:|||
Db  179 AEEVAFHAFTQWLFFRQWGALKAEAEALGIRIIGDMPIFVAEDSAEVWAHPEWFHLDEEG  238

Qy  241 NPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGF  300
         ||||||||| ||||||| |||||||||| ||: |:  :: :|:|:| : ||||||||||
Db  239 RPTVVAGVPPDYFSETGQRWGNPLYRWDVLEREGFSWIRRLEKALELFHLVRIDHFRGF   298

Qy  301 EAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGF  360
         |||| ||||| |||||||||||||||:  :: |: ::|| |||||||||||||:|| ||
Db  299 EAYWEIPASCPTAVEGRWVKAPGEKLFQKIQEVFGEVPVLAEDLGVITPEVEALRDRFGL  358

Qy  361 PGMKILQFAF-SGEDNAFLPHNYPAHGNVVVVSGTHDNDTTLGWFRTAPEAERAFMRAYL  419
         ||||:|||| ||::| |||||||||||:  :|||||||||||||||:||| ||||||
Db  359 PGMKVLQFAFDDGMENPFLPHNYPAHGRVVVVTGTHDNDTTLGWYRTATPHEKAFMARYL  418

Qy  420 ARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWRYA  479
         |:|||::||||||| ||| | | ||| ||: ||:|| |||:|:|||||  |||:||
Db  419 ADWGITFREEEVPWALMHLGMKSVARLAVYPVQDVLALGSEARMNYPGRPSGNWAWRLL   478

Qy  480 EGDLEPGLAAGLRALAEASQR  500
         ||||||||:||| |||:|
Db  479 PGELSPEHGARLRAMAEATER  499
```

RESULT 4
Q6JHX9_THEAQ
ID   Q6JHX9_THEAQ   PRELIMINARY;   PRT;   500 AA.
AC   Q6JHX9;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   4-alpha-glucanotransferase.
OS   Thermus aquaticus.
OC   Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OC   Thermus.
OX   NCBI_TaxID=271;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Park K.-H., Park J.-H.;
RL   Submitted (NOV-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY459352; AAR23242.1; -; Genomic_DNA.
DR   SMR; Q6JHX9; 1-499.
DR   GO; GO:0004134; F:4-alpha-glucanotransferase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR   InterPro; IPR003385; Glyco_hydro_77.
DR   Pfam; PF02446; Glyco_hydro_77; 1.
DR   TIGRFAMs; TIGR00217; malQ; 1.
KW   Transferase.
SQ   SEQUENCE   500 AA;   56973 MW; A03680E3B3113256 CRC64;

Query Match
Best Local Similarity 60.3%; Score 1635.5; DB 2; Length 500;
                       61.7%; Pred. No. 2.2e-115;
Matches 309; Conservative   59; Mismatches 130; Indels 3; Gaps 2;

RESULT 5

```
RESULT 5
Q6JHY0_THESC
ID  Q6JHY0_THESC PRELIMINARY;        PRT;   500 AA.
AC  Q6JHY0;
DT  05-JUL-2004 (TrEMBLrel. 27, Created)
DT  05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT  05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE  4-alpha-glucanotransferase.
OS  Thermus scotoductus.
OC  Bacteria; Deinococcus-Thermus; Thermales; Thermaceae;
OC  Thermus.
OX  NCBI_TaxID=37636;
RN  [1]
RP  NUCLEOTIDE SEQUENCE.
RA  Park K.-H., Park J.-H.;
RL  Submitted (NOV-2003) to the EMBL/GenBank/DDBJ databases.
DR  EMBL; AY459351; AAR23241.1; -; Genomic_DNA.
DR  SMR; Q6JHY0; 1-500.
DR  GO; GO:0004134; F:4-alpha-glucanotransferase activity; IEA.
DR  GO; GO:0016740; F:transferase activity; IEA.
DR  GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR  InterPro; IPR003385; Glyco_hydro_77.
DR  Pfam; PF02446; Glyco_hydro_77; 1.
DR  TIGRFAMs; TIGR00217; malQ; 1.
KW  Transferase.
SQ  SEQUENCE  500 AA;  57730 MW;  876B2DA81F61B33A CRC64;

Query Match            59.9%;
Best Local Similarity  60.3%;     Score 1626.5;  DB 2;  Length 500;
                                  Pred. No. 1.1e-114;
Matches 302;  Conservative 66;  Mismatches 130;  Indels 3;  Gaps 2;
```

```
Qy    1 MQLQRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQ  60
          |:| |||||||||  ||::|| ||  :: ||:|||:||||||| :::|||||||||||
Db    1 MELPRAFGLLLHPTSLPGPYGVGVLGEARDFLRFLKGAGGRFWQVLYQTRWPLLRRAFAGFRAR  60

Qy   61 SFSAFAGNPYLVDPEMLIEKGWLEQSEAPPYPTQRVDYGWLYQTRWPLLRRAFAGFRAR 120
          |||||||||::   :   :   ::|  :  |||:|||:|||| ::: |
Db   61 AFSAFAGNPYLIDLRLPLAEKGYLVLKD--PGFPQGRVDYGWLYAWKWPALKAAYQGFLEK 118

Qy  121 ASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAAREL 180
          :::   :   :||  :|||:||||||||| : :  |||.|||||::::: :::||
Db  119 APKEREDFPAFQKEASWLKDYALFMALKAQHGGLPWNRWPLLRRREEKAMKEAEAAL 178

Qy  181 AEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADG 240
          ||||| :.|:||||: |: ||| ::: :|||||||||||||:||::||:||::|:|::.
Db  179 AEEVAFHAWTQWLFFEAWKALKEAEALGIQIIGDMPIFVAEDSAEVWAHPEWFHLDEEG 238

Qy  241 NPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGF 300
          .|.||||||:||||||||.|||||||||:.||:. :|.|||:.::: |:||||:|||||
Db  239 RPLVVAGVPPDYFSETGQRWGNPLYRWDVLEREGFSFWIARLAKALELFHLVRDHFRGF 298

Qy  301 EAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGF 360
          |||||:|   :.|||||||||.||||| .:: : ..  |||||||||||||||:|||:|
Db  299 EAYWEIPASCPTAVEGRWVKASGEKLFDRIQEVPGEVPILAEDLGVITPEVEPLRDRYGL 358

Qy  361 PGMKILQFAFS-GEDNAFLPHNYPAHGNVVVVSGTHDNTTLGWFRTAPEAERAFMRAYL 419
          ||||:||||| .|:|.|||||||||||.||| ||||||||||||||| |.|||::|||
Db  359 PGMKVLQFAFDHGMENPFLPHNYPAHGRVVVYTGTHDNDTTLGWYRTASPHERAFLERYL 418

Qy  420 ARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWRYA 479
          | :||  ::::|| ||  | | ||.|:|||:| :|||.||.||||:|||  .|||||| :
Db  419 ADWGISFRQEEEVPWALMGLCLKSVARLAIYPVQDVLALGSEARMNYPGRPSGNWAWRLL 478

Qy  480 EGDLEPGLAAGLRALAEASQR 500
          .|:| .: :| | |:::|:.|
Db  479 PGQLTQEHAARLLAMAEATGR 499
```

RESULT 6

```
RESULT 6
Q4HAI6_9DEIO
ID  Q4HAI6_9DEIO PRELIMINARY;        PRT;   503 AA.
AC  Q4HAI6;
DT  13-SEP-2005 (TrEMBLrel. 31, Created)
DT  13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT  13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE  Glycoside hydrolase, family 77 (EC 2.4.1.25).
GN  ORFNames=DgeoDRAFT_1524;
OS  Deinococcus geothermalis DSM 11300.
OC  Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC  Deinococcaceae; Deinococcus.
OX  NCBI_TaxID=319795;
RN  [1]
RP  NUCLEOTIDE SEQUENCE.
RC  STRAIN=DSM 11300;
RG  US DOE Joint Genome Institute (JGI-PGF);
RA  Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,
RA  Hammon N., Israni S., Pitluck S., Richardson P.;
RT  "Sequencing of the draft genome assembly of Deinococcus geothermalis
RT  DSM 11300.";
RL  Submitted (MAY-2005) to the EMBL/GenBank/DDBJ databases.
RN  [2]
RP  NUCLEOTIDE SEQUENCE.
RC  STRAIN=DSM 11300;
RG  US DOE Joint Genome Institute (JGI-ORNL);
RA  Larimer F., Land M.;
RT  "Annotation of the draft genome assembly of Deinococcus geothermalis
RT  DSM 11300.";
RL  Submitted (JUN-2005) to the EMBL/GenBank/DDBJ databases.
CC  -!- CAUTION: The sequence shown here is derived from an
CC      EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC      preliminary data.
DR  EMBL; AAHE01000003; EAL83418.1; -; Genomic_DNA.
KW  Glycosyltransferase; Hydrolase; Transferase.
SQ  SEQUENCE  503 AA;  57157 MW;  7982B22BA8AFB5D5 CRC64;

Query Match            54.3%;
Best Local Similarity  57.7%;     Score 1475;  DB 2;  Length 503;
                                  Pred. No. 3.2e-103;
Matches 292;  Conservative 47;  Mismatches 155;  Indels 12;  Gaps 5;
```

```
Qy    1 MQLQRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQ  60
          |:| |||||||||  ||::|| ||  :: :| || |  ::||:||||||||:|||||||
Db    1 MELPRAYGLLLHPTSLPGPYGVGVLGEEAFAFLRFLGESSARYWQVLPLGPTGYGDSPYQ  60

Qy   61 SFSAFAGNPYLVDPEMLIEKGWLEQSEAPPYPTQRVDYGWLYQTRWPLLRRAFAGFRAR 120
          |||||||||:|  ::   :   : ::|    | :|| .||||| ||| ||:|: || ::
Db   61 SFSAFAGNPYLIDLRLRLRRRGYLRLED--PGFPBGRVVYGWLYAWKWPALREAFRGFQER 118

Qy  121 ASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAAREL 180
          ||::   : : || ::  |:||: |||||| :  | |||.||  |||::  ||: ||:|
Db  119 ASREEKEAFQPFGTRERSWLDDYTLFMALKPTHGGLPWNRWPMPLRLREEKALREASLAL 178

Qy  181 AEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADG 240
          ::.||::.|.||:|: . |: : :: ::| | |||||||||:||::||:||::|:|::|
Db  179 SQPVAFHAWTQWFFFRQWQALREAAEALGLSLIGDMPIFVAEDSAEVWAHPEWFHLDEEG 238

Qy  241 NPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGF 300
          .|||||||||:||||||||.|||||||||:.||:..:.|:||  ::|:||||||||||||
Db  239 RPTVVAGVPPDYFSETGQRWGNPLYRWDVLEREGFSWIERLRKALELFHLVRIDHFRGF 298

Qy  301 EAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGF 360
          |||||:|   ..||||||||||||||| .::: . ..|||||||||||||||||:||:|
Db  299 EAYWEIPASCPTAVEGRWVKAPGEKLFQKIQETFGRVPILAEDLGVITPEVEALRDRFGL 358

Qy  361 PGMKILQFAF-SGEDNAFLPHNYPAHGNVVVVSGTHDNDTTLGWFRTAPEAERAFMRAYL 419
          ||||:||||| .|:|.|||||||.||.||| ||||||||||||||| |.|||:.|||
Db  359 PGMKVLQFAFDDGMENPFLPHNYPSHGRVVVYTGTHDNDTTLGWYRTATPHERAFLGRYL 418

Qy  420 ARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWRYA 479
          | :||  ::::|: ||: | | ||.|:|| :| :|||.||.||||:|||  .||||| |:
Db  419 AEWGIGFQREEEIPWALMHLGMKSVARLAVYPVQDVLALGSEARMNYPGRPSGNWAWRLR 478

Qy  480 EGDLEPGLAAGLRALAEASQR 500
          .|:| .: |  |||:::|:.|
Db  479 PGQLLPEHGERLRWMAEATGR 499
```

```
Qy    1  MQLQRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQ   60
            : | :.|||||| :|||  : ||| | :|: | ||||||| :: |||||||.|||||||
Db    1  MLTKRSSGVLLHPTSLPGPYGIGELGAQARHFVDWLARAGQTTWQVMPLGPTGYGDSPYQ   60

Qy   61  SFSAFAGNPYLVDPEMLIEKGWLEQS--EAPPYPYTQRVDYGWLYQTRWPLLRRAFAGFR  118
            ::||||||| :||::|| ::| ||  :     |: | |||| |: : | :|| ::
Db   61  AFSAFAGNPYLIDLTTLREEGLLHASDFEGTPDFDAGRVDFGLQYVWRMQLGRAYAHFA  120

Qy  119  ARASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARARE  178
            || : ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  FGHHPELKAAFEAFKAEEAAWLDDYALFMALKDAHGGLPWNAWEPGTRDREPQALAAAQE  180

Qy  179  ELAEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAPDSSDVWANPQYFYLEA  238
            .|: : || :|| : ||| |: |: .| .:.| ||||| |:|||||:|| |:|||| :.
Db  181  QLAPNIERVKFIQFLFFRQWRALRTYARERGVGIGDIPIFVAMDSSDAWANREQFYFDE  240

Qy  239  DGNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFR  298
            ||||||||||||||||||||||||| ||||||||| ::|:|||||| :|| || |||||
Db  241  QGQPTVVAGVPPDYFSETGQLWGNPLYRWDVMEQDGFHWWIERPRGSLKLYDLIRIDHFR  300

Qy  299  GFEAYWEVPFGRNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGF  358
            || :|| |  ||||:   :  ||| |||||||:| | : .| |||||||:||||||| |
Db  301  GFAAYWEIPYPAENAIHGRWVPAPGHALLEAVRRALGQMPIIAEDLGVITPDVEQLRDDF  360

Qy  359  GFPGMKIIQFAFSGED---NAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFM  415
            |:||  :||||  |||   |||||||||||||:||||||||||| |||:|:||| ||:||
Db  361  GLPGMAVLHFAFGGGDFSVNAFLPHNLKA--NQVVYTGTHDNDTSRGHWQHADEQERQNF  418

Qy  416  RAYLARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLG-DNW  474
            || : || ||| | |||||| ||| |:|||||||| ||||  : :||||||||:||  ||
Db  419  RLYTHSDP----SEETFAWQLTEIALESRPNLAIVPLQDLLNLGSEARMNFPGTTGPHNW  474

Qy  475  AWRYAEGDLEPGLAAGLRALAEASQR  500
            ||||.|||||||||||||||| || |
Db  475  TWRYLAADLRPDLATKLRALTERTGR  500
```

RESULT 7
QSN4Q2_SYNP6
ID   QSN4Q2_SYNP6 PRELIMINARY;   PRT;   499 AA.
AC   Q5N4Q2;
DT   01-FEB-2005 (TrEMBLrel. 29, Created)
DT   01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT   01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE   4-alpha-glucanotransferase.
GN   OrderedLocusNames=syc0527_c;
OS   Synechococcus sp. (strain PCC 6301) (Anacystis nidulans).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX   NCBI_TaxID=269084;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=PCC6301;
RA   Sugita M.;
RT   "Complete genome structure of the unicellular cyanobacterium Anacystis
RT   nidulans 6301 (Synechococcus sp. PCC6301).";
RL   Submitted (DEC-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP008231; BAD78717.1; -; Genomic_DNA.
DR   GO; GO:0004134; F:4-alpha-glucanotransferase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR   InterPro; IPR003385; Glyco_hydro_77.
DR   Pfam; PF02446; TIGR00217; Glyco_hydro_77.
DR   TIGRFAMs; TIGR00217; malQ; 1.
KW   Complete proteome; Transferase.
SQ   SEQUENCE   499 AA; 56626 MW; 4A38325C7091FA0B CRC64;

Query Match   48.6%; Score 1318; DB 2; Length 499;
Best Local Similarity   51.6%; Pred. No. 2.6e-91;
Matches 256; Conservative 73; Mismatches 157; Indels 10; Gaps 5;

```
Qy    5  RAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQSFSA   64
            ::||||||||| ||:|||| :|| ||||||||:|| :||||:|||||||||:|||| :::
Db    4  RAAGILLHPTSLPGPFGSGDLGPASRQFLDWLATAGQLWQVLPLGPTGYGYSPYLCYSA   63
```

```
Qy   65  FAGNPYLVDPEMLIEKGWLEQSRAP--PPYPTQRVDYGWLYQTRWPLLRAFAGFRARAS  122
            ||||| |: |::|: |||:|| : :     || ||||| :|: :| :|| : :.|
Db   64  LAGNPALISPELLAEDGWLQESDWADCPAFPSDRVDFASVLPYRDQLLRRAYSQFLQRAA  123
```

```
Qy  123  AQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAREELAE  182
            :: :|||||||||||||||||||||||||||||||||||||||||||||| |:.:: | |
Db  124  SSDRLFQAFCEQEAHWLDDYALFMAIKLASQGQPWTEWPEALRQRQPQALAKARDRWGG  183
```

```
Qy  183  EVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAPDSSDVWANPQYFYLEAD-GN  241
            | :: :|:   |: |||| ::|:|: | | ||||| |:|||||:|| |:||: :.:. ::
Db  184  EIGFQQFLQWQFREQWLALREEAQARHISLIGDIPIYVAHDSADVWANPQFFALDPETGA  243
```

```
Qy  242  PTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGFE  301
            ||||:||  ||||||||||||| || |:    |  :  |:| :| | |:|||:|:| :
Db  244  VDQQAGVPPDYFSETGQLWGNPVVNWAALQADGYRWLQRLQQLLSLVDYIRIDHFRGLE  303
```

```
Qy  302  AYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGFP  361
            |: :||  | :.:||| | : ||||| :  :  : ||||:|||||||:|||||||||| |
Db  304  AFWSVPAGEETAIDGEWVKAPGADLLSTIRQKLGALPILAEDLGVITPEVEALRDRFELP  363
```

```
Qy  362  GMKILQFAP-SGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAYLA  420
            |||||||||   |||:| |||| |:||| | ||||:||||||||||:|:|||||:||||:
Db  364  GMKILQFAFDSGAGNAYLPHNYWGR--RWVAYTGTHDNDTTVGWFLSRNDSDRQTVLDYLG  422
```

```
Qy  421  RYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWRYAE  480
            |:|  | | |||||| ||| ||| ||||||||| | : ||||||||| :| || | :  :
Db  423  ----AESGWEIEWKLIRLAWSSTADWAIAPLQDVFGLDSSARMNRPGGATGNWDWRFSA  477
```

```
Qy  481  GDLEPGLAAGLRALAE  496
            |||:||||| ||||::
Db  478  DWLTGDRAQRLRRLSQ  493
```

RESULT 8
Q8YQG0_ANASP
ID   Q8YQG0_ANASP PRELIMINARY;   PRT;   502 AA.
AC   Q8YQG0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   4-alpha-glucanotransferase.
GN   OrderedLocusNames=alr3871;
OS   Anabaena sp. (strain PCC 7120).
OC   Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX   NCBI_TaxID=103690;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=21595285; PubMed=11759840;
RA   Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA   Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA   Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA   Yasuda M., Tabata S.;
RT   "Complete genomic sequence of the filamentous nitrogen-fixing
RT   cyanobacterium Anabaena sp. strain PCC 7120.";
RL   DNA Res. 8:205-213(2001).
DR   EMBL; BA000019; BAB75570.1; -; Genomic_DNA.
DR   PIR; AH2289; AH2289.
DR   HSSP; O87172; 1ESW.
DR   GO; GO:0004134; F:4-alpha-glucanotransferase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR   InterPro; IPR003385; Glyco_hydro_77.
DR   Pfam; PF02446; Glyco_hydro_77; 1.
DR   TIGRFAMs; TIGR00217; malQ; 1.
KW   Complete proteome; Transferase.
SQ   SEQUENCE   502 AA; 57349 MW; DD70A085DDC38DBE CRC64;

Query Match   48.1%; Score 1305.5; DB 2; Length 502;
Best Local Similarity   49.9%; Pred. No. 2.3e-90;
Matches 252; Conservative 80; Mismatches 162; Indels 11; Gaps 5;

```
Qy    1  MQLQRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQ   60
            |  | :|:||| |  |  |  ||| :  ||  :.:|||||||||||||
Db    1  MPFIRSSGVLLHPTSCPSRFGIGDLGLEAYKFIDFLEKSYQQYWQVLPLGPTGYGNSPYM   60

Qy   61  SFSAFAGNPYLVDPEMLIEKGWLEQSEAP--PPYPTQRVDYGWLYQTRWPLLRRAFAGFR  118
            :|| || |||:  |:||  |||| |        |  | ||||:|||||||| || |
Db   61  SYSALAGNHLLISPEKLLDEGLLSDEDFAHLPNFNEKVDFDQVAPIKIQLLKKACENFK   120

Qy  119  ARASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARARE  178
            ||| |  ||  | ||  |||||||||:|||||||||||| |:||| |:| | | |||
Db  121  TKASPLQQKGFAGFCETKSYWLDDYALFMALKDTKDSSSWHTWEPALAKREPDALEKVQR  180

Qy  179  ELAEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYL-E  237
            ||:||||||| |||||||||||| ||| ||||||||||||:|||||||||||||  :|| |
Db  181  QLTDEIFYKFIQYEFFRQWSELKSYANMRGIEIIGDIPIYVAHDSADVWANPDIFCLDE   240

Qy  238  ADGNPTVVAGVPRDYFSETGQLWGNPLYRWRWDVMERDNFAWCIARIRQSLKQCHLVRIDHF  297
             :|| | |||| |:|||| ||||||||||||||||:|||||:||:|||||||||:| ||||
Db  241  ETGEVALMAGVPPDYFSATGQLWGNPVVNWEELQKQDFKWWVQRFEAMLDYVDVIRIDHF   300

Qy  298  RGFEAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDG  357
            |||:|| ||| |||:  || |||| :|  |   |   | |||||:| || |:| ||| |
Db  301  RGFEAFWTVPQGGETAMNGEWVTAPGEELFDRAIKQKLGKLPVLAEDLGVITPEVEALRDK   360

Qy  358  FGFPGMKILQFAF-SGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAEARAFMR  416
            ||||||:||||| |  |:|| ||| |  |:  :  |  |||||||||||| |: | || |
Db  361  YEFPGMKVLQFAFGSDPGNPFLPFNYSR--NFVVYTGTHDNDTTVGWFNTANDYEKNNLL   418

Qy  417  AYLARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGPEARMNFPGRLGDNWAW   476
            ||| ||:|||||||| |:||||||||| ||  : | |  |   | | |  |||:| |
Db  419  LYLG-----CISPEGIHWDLIRLALSSVANQAIVPLQDILGLGNEARMNFPSIAEGNWAW   473

Qy  477  RYAEGDLEPGLAAGLRALAEEASQRA   501
            ||  :  : || | |:|: |  |
Db  474  RYDFAVLTDELSDRLKILTTKLYGRA   498
```

Matches 252; Conservative 73; Mismatches 155; Indels 32; Gaps 10;

```
Qy    1  MQLQRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQ   60
            | | :|:|||  |  |  ||| :  ||  :.:||||||||||||:
Db    1  MLLPRASGILLHPTSLFGPYGIGSLGSEAREFIDLMAITGQRFWQILPLLAPTGYGNCPYM   60

Qy   61  SFSAFAGNPYLVDPEMLIEKGWLEQSEAPPP---------YPTQRVDYGWLYQTRWPLLR  111
            :|||||||| |||| :|:|||||||:    |         : || | |||||:|||||||
Db   61  SCSAFAGNPLLVDPEDLVQALIDKDDLAPDELKTTAGWHHPKINFEEAVAYKTR--LLE   118

Qy  112  RAFAGFRARASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPA  171
            : | ||  :  .:   |:|| | ||||:|||||||||||  ::| ||::|| |  |||:
Db  119  TACENFAAKGDS-DKKLFWAYCSREGWMLDDYALFMALKEANDTLPWQEWDQGLALREGA  177

Qy  172  ALARARELAEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANP  231
            .| :  :||:: | | | ::||||: |::  ||  : |: || |:||::  | ||||:|
Db  178  VLKQQSRELRDRVFYHQFSQYLFSRQWSALKQYANGRGVSILGDLPIYVAADSADVWAHP  237

Qy  232  QYFYLEADGNPTVVAGVPRDYFSETGQLWGNPLYRWRWDVMERDNFAWCIARIRQSLKQCHL  291
            : : |:|:|||||||||| ||||:|||||||:|  | . |||| :||  :.|||:::| |
Db  238  HLFQLDESGNPTVVAGVPPDYFSATGQLWGNPIYDWEAMERNGYAWWINRIKRMLDMVDL  297

Qy  292  VRIDHFGFEAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEV  351
            ||||||  :|:||| ||:   | ||| |:||| |||   :.|  | ::|||||||:||:|
Db  298  VRIDHFRCFASYWEVEAGRTTAIDGRWIKGPGPQLFDAIEKALGSLPVVAEDLGVLTDEV  357

Qy  352  EALRDGFGFPGMKILQFAPSG-EDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAP--  408
            | ||| :| |||||:||||.|  :| :|||||  ::|  : |.||||||||||::: .
Db  358  ETLRDRYGLFGMKVLQFAPDGMANNIHLPHNHVR--SSIVYTGTHDNDTTVGWYNHSPRW  415

Qy  409  --EAERAFMRAYLARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNF  466
              :|  | :  |:|  :   ||| :|  |:| |. :: | |::::|||||| |.||||:
Db  416  VHEATERYL--YLN------TGWEIHWALVRCALASVCNLSIPMQDVLGLGSEARMNT  498

Qy  467  PGRLGDNWAWRYAEGD-LEPGL---AAGLRAL   494
            || .| |||:| :   :  |.|    |||:|
Db  467  PGVAEGNWDWRVGTFEAVEPWMRERLADLTAL   498
```

RESULT 9
Q7NG57_GLOVI
ID   Q7NG57_GLOVI PRELIMINARY;   PRT;   504 AA.
AC   Q7NG57;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   4-alpha-glucanotransferase.
GN   OrderedLocusNames=gll3316;
OS   Gloeobacter violaceus.
OC   Bacteria; Cyanobacteria; Gloeobacteria; Gloeobacterales; Gloeobacter.
OX   NCBI_TaxID=33072;
RN   [1]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC   STRAIN=PCC 7421;
RX   MEDLINE=22977040; PubMed=14621292;
RA   Nakamura Y., Kaneko T., Sato S., Mimuro M., Miyashita H., Tsuchiya T.,
RA   Sasamoto S., Watanabe A., Kawashima K., Kishida Y., Kiyokawa C.,
RA   Kohara M., Matsumoto M., Matsuno A., Nakazaki N., Shimpo S.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of Gloeobacter violaceus PCC 7421, a
RT   cyanobacterium that lacks thylakoids.";
RL   DNA Res. 10:137-145(2003).
DR   EMBL; BA000045; BAC91257.1; -; Genomic_DNA.
DR   HSSP; O87172; 1ESW.
DR   GO; GO:0004134; F:4-alpha-glucanotransferase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR   InterPro; IPR003385; Glyco_hydro_77.
DR   Pfam; PF02446; Glyco_hydro_77; 1.
DR   TIGRFAMs; TIGR00217; malQ; 1.
KW   Complete proteome; Transferase.
SQ   SEQUENCE   504 AA; 57041 MW; B781C4C1AFD23BE2 CRC64;

Query Match   46.0%; Score 1248; DB 2; Length 504;
Best Local Similarity   49.2%; Pred. No. 5.2e-86;

RESULT 10
Q74DY3_GEOSL
ID   Q74DY3_GEOSL PRELIMINARY;   PRT;   493 AA.
AC   Q74DY3;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   4-alpha-glucanotransferase (EC 2.4.1.25).
GN   Name=malQ; OrderedLocusNames=GSU1182;
OS   Geobacter sulfurreducens.
OC   Bacteria; Proteobacteria; Deltaproteobacteria; Desulfuromonadales;
OC   Geobacteraceae; Geobacter.
OX   NCBI_TaxID=35554;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=PCA / ATCC 51573;
RX   PubMed=14671304; DOI=10.1126/science.1088727;
RA   Methe B.A., Nelson K.E., Eisen J.A., Paulsen I.T., Nelson W.C.,
RA   Heidelberg J.F., Wu D., Wu M., Ward N.L., Beanan M.J., Dodson R.J.,
RA   Madupu R., Brinkac L.M., Daugherty S.C., DeBoy R.T., Durkin A.S.,
RA   Gwinn M.L., Kolonay J.F., Sullivan S.A., Haft D.H., Selengut J.,
RA   Davidsen T.M., Zafar N., White O., Tran B., Romero C., Forberger H.A.,
RA   Weidman J.F., Khouri H.M., Feldblyum T.V., Utterback T.R.,
RA   Van Aken S.E., Lovley D.R., Fraser C.M.;
RT   "Genome of Geobacter sulfurreducens: metal reduction in subsurface
RT   environments ";
RL   Science 302:1967-1969(2003).
DR   EMBL; AB017180; AAR34558.1; -; Genomic_DNA.
DR   TIGR; GSU1182; -.
DR   GO; GO:0004134; F:4-alpha-glucanotransferase activity; IEA.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR   GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR   InterPro; IPR003385; Glyco_hydro_77.
DR   Pfam; PF02446; Glyco_hydro_77; 1.

```
DR   TIGRFAMs; TIGR00217; malQ; 1.
KW   Complete proteome; Glycosyltransferase; Transferase.
SQ   SEQUENCE   493 AA;  55272 MW;  60ED64942D575C2F CRC64;

Query Match          45.9%;    Score 1245;  DB 2;  Length 493;
Best Local Similarity  52.3%;    Pred. No. 8.6e-86;
Matches 253;  Conservative  49;  Mismatches 172;  Indels  10;  Gaps  4;

Qy    1 MQLQRAFGILLHPTSFPGRWIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQ  60
          |: |:||||||||||| |||||||||||||||||  |: ||||  |||:|:||:|:|||
Db    1 MRQKRRSGILLHPTSLPGPGGIGSLGKECRHFIDFLEEAGQSLWQVLPLGPAAYGNSPYS  60

Qy   61 SFSAFAGNPYLVDPEMLIEKGWLEQSEAPPPYPTQRVDYGWLYQTRWPLLRRAFAGFRAR 120
          |||||||||| ::| :|||: :||||:|| ||||||||||||| :|| ::||||:||||
Db   61 CYSAFAGNPLLIDLATLVDEGDLEPEEAVAEVSADHVDFPAVDAFKSRALRTAAARFHAH 120

Qy  121 ASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAREEL 180
          ||:||||||||||||||:| ||||||||||:||  :|::||||||||||||||||||
Db  121 GDTDRKQEFWHFCDTTP-WLHDAALFMALKEHFGGKSNNTWPKEIARRESAALEKYSVRL 179

Qy  181 AEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADG 240
          |:::||:  :||||||||||||||||||||||||||||||||:|:|| ||||||||: |
Db  180 GVAIGEHKYIQWQFSRQWRHVREYANSKGIAVVGDIPIFVAFDSTDVWTNPGLFKLDDKG 239

Qy  241 NPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGF 300
          ||||||||||||||||||||||||||||  || |||||||:|||||  ||||||||||||
Db  240 KPTVVAGVPPDYFSKTGQRWGNPLYDWDAMAADGFRWNIERFRGSFALHDVVRVDHFRGF 299

Qy  301 EAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGF 360
          |: ||| || | ||: | |: ::||| ||||||||||||||||||||||||||: ||| |
Db  300 EACWEVPAKEKTAVNGQWVPAPGTALFDAVIGAVGSLPIIAAEDLGVITPAVEELRDRYRF 359

Qy  361 PGMKILQFAF-SGEDNAFLPHNYPAHGNVVVYSGTHDNTTLGWFRTAPEAERAFMRAYL 419
          ||||||||||  || :|||:||||||| ||||||||||||:|||||:|||||||||||:|
Db  360 PGMKILQFAFDSGPGNPYLPHNFTR--DCVVVTGTHDNDTTCGWFQSLSPRERREVLAYT 417

Qy  420 ARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWRYA 479
          ||:|||||||||||| ||| |||||||| :||::||| ||||| |||||||||: |:
Db  418 DTTG-----KEVHWELIRLGMASVADMAIFPLQDVMGLDGGSSRMNLPGTPRGNWSWRYS 471-

Qy  480 EGDL 483
          || |
Db  472 EGML 475
```

RESULT 11
MALQ_SYNY3
```
ID   MALQ_SYNY3      STANDARD;      PRT;   505 AA.
AC   P72785;
DT   30-MAY-2000 (Rel. 39,  Created)
DT   30-MAY-2000 (Rel. 39,  Last sequence update)
DT   10-MAY-2005 (Rel. 47,  Last annotation update)
DE   4-alpha-glucanotransferase (EC 2.4.1.25) (Amylomaltase)
DE   (Disproportionating enzyme) (D-enzyme).
GN   Name=malQ; OrderedLocusNames=sll1676;
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
RN   [1]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX   MEDLINE=97061201; PubMed=8905231;
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA   Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA   Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
CC   -!- CATALYTIC ACTIVITY: Transfers a segment of a 1,4-alpha-D-glucan to
CC       a new position in an acceptor, which may be glucose or a 1,4-
CC       alpha-D-glucan.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the disproportionating enzyme family.
CC   ------------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ------------------------------------------------------------------------
DR   EMBL; BA000022; BAA16800.1; -; Genomic_DNA.
DR   PIR; S74648; S74648.
DR   HSSP; O87172; 1ESW.
DR   InterPro; IPR003385; Glyco_hydro_77.
DR   Pfam; PF02446; Glyco_hydro_77; 1.
DR   TIGRFAMs; TIGR00217; malQ; 1.
KW   Carbohydrate metabolism; Complete proteome; Glycosyltransferase;
KW   Transferase.
SQ   SEQUENCE   505 AA;  57180 MW;  488623C9FA4C9D59 CRC64;

Query Match          45.0%;    Score 1222.5;  DB 1;  Length 505;
Best Local Similarity  48.6%;    Pred. No. 4.5e-84;
Matches 242;  Conservative  76;  Mismatches 169;  Indels  11;  Gaps  6;

Qy    4 QRAFGILLHPTSFPGRWIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQSFS  63
          :|||||||||||||||||||||||||||||||||||:|||||||||||||: ||:|:||
Db    4 KRCSGILLHPTSLPSRFGIGDLGDGAFQPIDFLADADQSVWQILPLGPTGFGNSPYLCYS  63

Qy   64 AFAGNPYLVDPEMLIEKGWLEQS--EAPPPYPTQRVDYGWLYQTRWPLLRRAFAGFRARA 121
          |:|||| |:||||:::||: |||  :|| || |  |||:|||||  | | | ||:|:|
Db   64 ALAINPWLISLDRLAEEGPLPPSLLDQAPPFTNPRVDYDQAIAYKSQVLKQAFAQFRTNI 123

Qy  122 SAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAREELA 181
          ||||||| |||||||||| |||:||:|| |  |:  |:|: :| : :||| ||  || |
Db  124 ELAIEQEFAEFCQAQSDWLADYALFMAIKEAHNGAGWHQWDKDIAWREPEALKIWGDRLK 183

Qy  182 EEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEAD-G 240
          ||||| ||||||||||| ||||||:|||||:||||||||||||:|:|||||||||||   |
Db  184 TEVLYHQFLQFLGFRQWQEVKAYANQRHIAIFGDLPIYVAHDSADVWANPENFCLDPETG 243

Qy  241 NPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGF 300
          ||||||||||||||||| ||||||||:|| ||||| ||||||:|||  ::|||||||||
Db  244 EAAMMAGVPPDYFSATGQLWGNPVYDWETLKATGFAWWIKRFKANLQYLDIVRIDHFRGF 303

Qy  301 EAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSD-APIIAEDLGVITPEVEALRDGFG 359
          |: ||||| |  | |: |  ||| ||  |:|||| |   |||:|||||||||:||| |
Db  304 ESYWGVPQGEKTAENGBWYPAPGKEFFQALGKALGDNLPIVADEAGVGVITPEVEALRDEFN 363

Qy  360 FPGMKILQFAF-SGEDNAFLPHNYPAHGNVVVYSGTHDNTTLGWFRTAPEABRAFMRAY 418
          ||||||||||  || :|||:||||:||||  ||||||||||:|||||:|||::|:|||||
Db  364 FPGMKVLHFAFDSDRGNPFLPNY-SNGNAVVYGTHDNDTTVGWFQERSEDDQKVINY 422

Qy  419 LARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWRY 478
          | :||||||||||||| |||| |||:| :||:::|| ||| ||||||||||:|| |||
Db  423 LG------CVCNEGIHWSLIRLASSSVAALAIFPLQDIIGLGSDCRMNLPGTAAGNWGWRY 477

Qy  479 AEGDLEPGLAAGLRALAE 496
          || || || | |:|
Db  478 HPDQLNDWLSGHLSFITE 495
```

RESULT 12
Q4NUB1_9DELT
```
ID   Q4NUB1_9DELT  PRELIMINARY;      PRT;   518 AA.
AC   Q4NUB1;
DT   13-SEP-2005 (TrEMBLrel. 31,  Created)
DT   13-SEP-2005 (TrEMBLrel. 31,  Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31,  Last annotation update)
DE   Glycoside hydrolase, family 77 (EC 2.4.1.25).
GN   ORFNames=AdehDRAFT_2695;
OS   Anaeromyxobacter dehalogenans 2CP-C.
OC   Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC   Cystobacterineae; Myxococcaceae; Anaeromyxobacter.
OX   NCBI_TaxID=290397;
```

RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=2CP-C;
RG   US DOE Joint Genome Institute (JGI-PGF);
RA   Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,
RA   Hammon N., Israni S., Pitluck S., Richardson P.;
RT   "Sequencing of the draft genome assembly of Anaeromyxobacter
RT   dehalogenans 2CP-C.";
RL   Submitted (MAY-2005) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=2CP-C;
RG   US DOE Joint Genome Institute (JGI-ORNL);
RA   Larimer F., Land M.;
RT   "Annotation of the draft genome assembly of Anaeromyxobacter
RT   dehalogenans 2CP-C.";
RL   Submitted (JUN-2005) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC   EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC   preliminary data.
DR   EMBL; AAHD01000014; EAL79233.1; -; Genomic_DNA.
KW   Glycosyltransferase; Hydrolase; Transferase.
SQ   SEQUENCE 518 AA; 56400 MW; A6CF058B1AB9C586 CRC64;

Query Match   45.0%; Score 1221; DB 2; Length 518;
Best Local Similarity   49.8%; Pred. No. 6e-84;
Matches   255; Conservative   53; Mismatches 180; Indels 24; Gaps   7;

Qy   5 RAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQSFSA   64
           .:|||:||||||:.:|||||| ||||||:|||||||:: : :||||||:|||||:|| :
Db   4 RRSGLLLHPTSLPGPHGAGDLGAAAHRFAGWLADAGQRLWQVLPLGPTGFGDSPYQALSS   63

Qy   65 FAGNPYLVDPEMLIEKGWLEQSE---APPPYPTQRVDYGWLYQTRWPLLRRAFAGFRARA 121
           ||||||| | || ||||||||: || ||: || || : |||:|||:|| |:|: ||
Db   64 HAGNPLLVSLEVLRNEGWLDDDADLSGAPAGEP-GRADLHAALEWKRERLARAARAFRAGA 122

Qy   122 SAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARARRELA 181
           |||||| |||||||||:|||| :||:|||| |||||| |||||||||: ||| : ||:|
Db   123 DGERAAELEDFRAREAGWLEDWALFALLKAAHGGRPWTAWPAPLARRERAALESARARFA 182

Qy   182 EEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADGN 241
           :||::|: : :|||::|| | | :||| |||| :||| |||| | :|||:|||:||||
Db   183 HEVFAEVFAQFAFHRQWAALRARCRALGIELMGDLPLFIYVAHDSAEVWARPELFRLDDAGE 242

Qy   242 PTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGFE 301
           || ||||||||||| ||||||||| ||:||:||| | : ||| : :|:|||:||||||
Db   243 PAAVAGVPPDYFSATGQRWGNPLYDWEAVAREGWRFWIERVRGTLALVDRIRLDHFRGFE 302

Qy   302 AYWEVPPGRPNAVEGRWKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGFP 361
           |||:|| ||||||||||| |||||||||:|| : ||||||||||||||||||||:|| |
Db   303 AYWEIPAGAPTAERGRWVPGPGARLFEALLRALGPLPFVAENLGVITPEVEALRRRFGLP 362

Qy   362 GMKILQFAFSGEDNA--FLPHNYPAHGNVVVYSGTHDNDTTLGWF------RTAPEA- 410
           ||||||||| | : :: | ||| || : || ||||||||||| || ||
Db   363 GMAILQFAFGDDPQAPTFQPHNYAP--DLVAYTGTHDNDTVAGWWGGGAGDSVRTAEAVA 420

Qy   411 -ERAFMRAYLARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGR 469
            :|| |||||||||||||||||||||||||||||||||||| |||:||||| |||||| |
Db   421 REKAFALEYLGVDG------RDVPGAMIRAVLASVADTAVFPLQDALGLGSEARMNTPAT 474

Qy   470 LGDNWANRYAEGDLEPGLAAGLRALALARASQRA   501
           ||| |:|||||||| |:||||||||:::|| |
Db   475 LGGNWRNRVREEALDAALAARLRRLAAVYGRA   506

RESULT 13
Q8DKZ1_SYNEL
ID   Q8DKZ1_SYNEL PRELIMINARY;   PRT; 518 AA.
AC   Q8DKZ1;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   4-alpha-glucanotransferase.

GN   OrderedLocusNames=tlr0708;
OS   Synechococcus elongatus (Thermosynechococcus elongatus).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX   NCBI_TaxID=32046;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=BP-1;
RX   MEDLINE=22225144; PubMed=12240834;
RA   Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA   Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA   Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA   Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the thermophilic cyanobacterium
RT   Thermosynechococcus elongatus BP-1.";
RL   DNA Res. 9:123-130(2002).
DR   EMBL; BA000039; BAC08259.1; -; Genomic_DNA.
DR   HSSP; O87172; 1ESW.
DR   GO; GO:0004134; F:4-alpha-glucanotransferase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR   InterPro; IPR003385; Glyco_hydro_77.
DR   Pfam; PF02446; Glyco_hydro_77; 1.
DR   TIGRFAMs; TIGR00217; malQ; 1.
KW   Complete proteome; Transferase.
SQ   SEQUENCE 518 AA; 58347 MW; 3DA8B8A5DF2474B1 CRC64;

Query Match   44.8%; Score 1217; DB 2; Length 518;
Best Local Similarity   49.8%; Pred. No. 1.2e-83;
Matches   249; Conservative   61; Mismatches 174; Indels 16; Gaps   6;

Qy   1 MQLQRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQ   60
           | |||:|||||||:.:||||||||||||||:|||||||||||||||||| ||:||:|:|
Db   1 MHFPRSCGLLLHPTSLPGGHGGHGIGELGTAAREFLEFLVASDQQYWQVLPLGPTGFGNSPYM   60

Qy   61 SFSAFAGNPYLVDPEMLIEKGWLEQSEAPPPYPT----QRVDYGWLYQTRWPLLRRAFAG 116
           .:||:||||| | || |||||||:|||||||| :|||| || :|||:|||:|| |:
Db   61 CYSAMAGNPLLISLEEVAKAGWL--TEADLGQITLENRDRVDFDAVISQKLPLLRLLAAQR 118

Qy   117 FRARASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARA 176
           ||||||||||||||||||||:|||| :||:|||| |||||| |||||||||: |||||||
Db   119 FQSQATPGDWQAFRDFQALAHYWLPNYALFMAIKDHHQGQPWYEWPAPLRDREPTALALAI 178

Qy   177 REELAEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYL 236
           |||:| : :||:|||||||:|| | |: ||| |||| :||| |||| | :|||:|||
Db   179 QVVLKDRIFEYEFQQFLFYQQWHALKEAANQQGIQIIIGDIPIYVAHDSVDVWAFPQFFEL 238

Qy   237 EADGNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRID 295
           || ||| ||||||||||| ||||||||| ||:||:||| |||||||| : :|:|||:|
Db   239 NPETGAAALMAGVPPDYFSATGQLWGNPIYNWKALAADGYSWWIERFRALLAAYVDIIRVD 298

Qy   296 HFRGFEAYWEVPPGRPNAVEGRWKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALR 355
           ||||||||| || ||||||||||| |||||||||:|| :||||||||||||||||||||
Db   299 HFRGFQAYWQVPEGEKTAVNGQWQPGPGAAFFEALQAALGRLPILAEDLGDITPDVVIALR 358

Qy   356 DGFGFPGMKILQFAF-SGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAF 414
           |:|| ||||||||| |:|||||||||||:|| ||| |||||||| ||:|| ||||
Db   359 DQFQFPGMKILQFAFGGGSDNPFLPFN--QERNCVVTGTHDNDTTVGWYRNLSNWEROR 416

Qy   415 MRAYLARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNW 474
           ||||||||||||||||||||||||||||||||:|:||:|| ||| :||||||| ||||
Db   417 FIDYLG------YTPSEPHWALIRMALGTVANQAIIPVQDLLGLDSHARMNFPGTDQGNW 470

Qy   475 AWRYAEGDLEPGLAAGLRAL   494
           |||:|||||| ||||||||:
Db   471 AWRLTPGQLTPELAAAHLRHL   490

RESULT 14
MALQ_STRPN
ID   MALQ_STRPN   STANDARD;   PRT; 505 AA.
AC   P0A3Q0; P29851;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)

```
Qy  240 GNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRG 299
            |||  ||| |  | |||||||||| |  |      ||  |  | ||     ||||||||||
Db  240 GKATCIAGCPPDEFSVTGQLWGNPIYDWEAMDKDGYKWWIERLRESFKIYDIVRIDHFRG 299

Qy  300 FEAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFG 359
            || |||  |     ||| ||| || |||||        |||||||| |  | | |  | |
Db  300 FESYWEIPAGSDTAAPGEWVKGPGYKLFAAVKEELGELNIIAEDLGFWTDEVIELRERTG 359

Qy  360 PPGMKIIQPAFSGEDNAP-LPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAY 418
            ||||||| || |    ||           || | |||||| | || | |        | |
Db  360 PPGMKIIQFAFNPEDESIDSPHLAPA--NSVMYTGTHDNNTVLGWYRN--EIDDA-TREY 414

Qy  419 LARYGIRCLSEYE-VAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWR 477
            :||| |      |      | | |          |||  | |   |||||| ||| ||:||
Db  415 MARYTNR--KEYETVVHAMLRTVFSSVSFMAIATMQDILELDEAARMNFPSTLGGNWSWR 472

Qy  478 YAEGDLEPGLAAGLRALAEASQR 500
               | |     | |          |
Db  473 MTEDQITPAVEEGLLDLTTIYRR 495
```

RESULT 15

MALQ_STRR6

```
ID   MALQ_STRR6     STANDARD;      PRT;   505 AA.
AC   P0A3Q1; P29851;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   4-alpha-glucanotransferase (EC 2.4.1.25) (Amylomaltase)
DE   (Disproportionating enzyme) (D-enzyme).
GN   Name=malQ; Synonyms=malM; OrderedLocusNames=spr1917;
OS   Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=171101;
RN   [1]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX   MEDLINE=21429245; PubMed=11544234;
RX   DOI=10.1128/JB.183.19.5709-5717.2001;
RA   Hoskins J., Alborn W.E. Jr., Arnold J., Blaszczak L.C., Burgett S.,
RA   DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA   Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA   LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA   McAhren S.M., McHenney M., McLeaster K., Mundy C.W., Nicas T.I.,
RA   Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA   Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA   Zook C.A., Baltz R.H., Jaskunas S.R., Rosteck P.R. Jr., Skatrud P.L.,
RA   Glass J.I.;
RT   "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL   J. Bacteriol. 183:5709-5717(2001).
CC   -!- CATALYTIC ACTIVITY: Transfers a segment of a 1,4-alpha-D-glucan to
CC       a new position in an acceptor, which may be glucose or a 1,4-
CC       alpha-D-glucan.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to the disproportionating enzyme family.
CC   -----------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   -----------------------------------------------------------------------
DR   EMBL; AE008556; AAL00719.1; -; Genomic_DNA.
DR   PIR; B98111; B98111.
DR   HSSP; Q87172; 1ESW.
DR   InterPro; IPR003385; Glyco_hydro_77.
DR   Pfam; PF02446; Glyco_hydro_77; 1.
DR   TIGRFAMs; TIGR00217; malQ; 1.
KW   Carbohydrate metabolism; Complete proteome; Glycosyltransferase;
KW   Transferase.
SQ   SEQUENCE   505 AA;  58076 MW;  D4529A000A6D01A5 CRC64;
```

```
Query Match          44.0%;  Score 1194;  DB 1;  Length 505;
Best Local Similarity  48.1%;  Pred. No. 6.5e-82;
Matches  242;  Conservative  77;  Mismatches  168;  Indels  16;  Gaps  9;

Qy    4 QRAFGILLHFTSFPGRWGIGALGREAERFLDWLADAGARWQVLPLGPTSYGDSPYQSFS 63
          || |  | |: ||| |||| :|  : : |:|:|     |:|||| ||||||||||||||
Db    3 KRQSGVLMHISSLPGAYGIGSFGQSAYDFVDFLVRTKQRYWQILPLGATSYGDSPYQSFS 62

Qy   64 AFAGNPYLVDPEMLIEKGWLEQSEAPP---PYPTQRVDYGWLYQTRWPLLRRAFAGFRAR 120
          |||||  ::| ::|:: | ||: :        :::::VDY  :Y  RWPLL :|  :|  
Db   63 AFAGNTHFIDLDILVEQGLLEASDLEGVDFGSDASEVDYAKIYYARRPLLEKAVKRFFEV 122

Qy  121 ASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEW-SPELRDREPAALARAREE 179
           ...: :  : | : :  ||| :| :|| | :|| : ||:| |  |    :  |:|:: :
Db  123 GDVKD---FEKFAQDNQSWLELFAEYMAIKEYFDNLAWTEWPDADARARKASALESYREQ 179

Qy  180 LAEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEAD 239
          || : | |  : : ::: |   || |   :  :IGDMPI:VA DSSD WANP  | :  :
Db  180 LADKLVYHRVTQYFFQQWLKLKAYANDNHIEIVGDMPIYVAEDSSDMWANPHLFKTDVN 239
```

```
Query Match                    44.0%;   Score 1194;   DB 1;   Length 505;
Best Local Similarity  48.1%;   Pred. No. 6.5e-82;
Matches  242;   Conservative  77;   Mismatches 168;   Indels  16;   Gaps   9;

Qy    4 QRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQSFS  63
        :|:||:|:.|:.|:::|||::|.|:|:::|       |::||:||||||||||||||||
Db    3 KRQSGVLMHISSLPGAYGIGSFGQSAYDFVDFLVRTKQRYWQILPLGATSYGDSPYQSFS  62

Qy   64 AFAGNPYLVDPEMLIEKGWLEQSEAPP---PYPTQRVDYGWLYQTRWPLLRRAPAGFRAR 120
        ||||| :.|:|: :||:::|:|.: :      ||:  :  : |:|||||::|||||||
Db   63 AFAGNTHFIDLDILVEQGLLEASDLEGVDFGSDASEVDYAKIYYARRPLLEKAVKRFFEV 122

Qy  121 ASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEW-SPELRDREPAALARAREE 179
        |     :  : ||:    :|::: ||:    :  : ::::||:.|:||:: :|||:::: :
Db  123 GDVKD---FEKFAQDNQSWLELFAEYMAIKEYFDNLAWTEWPDADARARKASALESYREQ 179

Qy  180 LAEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEAD 239
        |:::::::|||:|.:: :|:::: :||.|:|:::: |||:||:|:|||||||.|||:.:
Db  180 LADKLVYHRVTQYFFFQQWLKLKAYANDNHIEIVGDMPIYVAEDSSDMWANPHLFKTDVN 239

Qy  240 GNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRG 299
        |::|.:|:|||||::|:||||||||||||||::::::|||:||:|::|:::||||||||||
Db  240 GKATCIAGCPPDEFSVTGQLWGNPIYDWEAMDKDGYKWWIERLRESFKIYDIVRIDHFRG 299

Qy  300 FEAYWBVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFG 359
        ||:|::||:|:||||:| |||||:| :   :|:: :: :::||||:: :|||:|:||:|:
Db  300 FESYWEIPAGSDTAAPGEWVKGPGYKLFAAVKEELGELNIIAEDLGFMTDEVIELRERTG 359

Qy  360 FPGMKILQFAFSGEDNAF-LPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAY 418
        |||||||||||.:|::|  ||:|| |:|||.::||||||:::||||||::|::|::|||
Db  360 FPGMKILQFAFNPEDESIDSPHLAPA--NSVMYTGTHDNNTVLGWYRN--EIDDA-TREY 414

Qy  419 LARYGIRCLSEYE-VAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWR 477
        |||||::|:||:: :::|:|::|:||:| ::: |: || |:||:|||||||:||||:||
Db  415 MARYTNR--KEYETVVHAMLRTVFSSVSFMAIATMQDLLELDEAARMNFPSTLGGNWSWR 472

Qy  478 YAEGDLEPGLAAGLRALAEASQR 500
        |::|||:||:::|:|::: ::|
Db  473 MTEDQLTPAVEEGLLDLTTIYRR 495


Search completed: January 10, 2006, 14:12:38
Job time : 167 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2006, 14:06:43 ; Search time 47 Seconds
                                        (without alignments)
                                        881.288 Million cell updates/sec

Title: US-10-618-976-2
Perfect score: 2714
Sequence: 1 MQLQRAFGILLHPTSFPGRW.........DLEPGLAAGLRALAEASQRA 501

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_AA:*
   1: /cgn2_6/ptcdata/1/iaa/S_COMB.pep:*
   2: /cgn2_6/ptcdata/1/iaa/6_COMB.pep:*
   3: /cgn2_6/ptcdata/1/iaa/H_COMB.pep:*
   4: /cgn2_6/ptcdata/1/iaa/PCTUS_COMB.pep:*
   5: /cgn2_6/ptcdata/1/iaa/RE_COMB.pep:*
   6: /cgn2_6/ptcdata/1/iaa/backfilesl.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the total score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2714	100.0	501	2	US-09-687-360-2	Sequence 2, Appli
2	1196	44.1	505	2	US-09-583-110-4228	Sequence 4228, Ap
3	1076	39.6	458	2	US-09-107-433-2939	Sequence 2939, Ap
4	777.5	28.6	298	1	US-08-838-543-5	Sequence 5, Appli
5	694	25.6	291	1	US-08-838-543-6	Sequence 6, Appli
6	606.5	22.3	470	2	US-09-902-540-9896	Sequence 9896, Ap
7	451	16.6	526	2	US-09-198-452A-347	Sequence 347, App
8	451	16.6	536	2	US-09-438-185A-328	Sequence 328, App
9	299	11.0	681	2	US-09-252-991A-24629	Sequence 24629, A
10	278	10.2	743	2	US-09-489-039A-11073	Sequence 11073, A
11	273.5	10.1	496	2	US-08-838-543-2	Sequence 2, Appli
12	197	7.3	323	1	US-08-838-543-4	Sequence 4, Appli
13	187	6.9	330	1	US-08-838-543-3	Sequence 3, Appli
14	123	4.5	1332	2	US-09-252-991A-25772	Sequence 25772, A
15	105.5	3.9	508	2	US-09-583-110-5034	Sequence 5034, Ap
16	105.5	3.9	594	2	US-09-107-433-4327	Sequence 4327, Ap
17	104	3.8	436	2	US-09-252-991A-17273	Sequence 17273, A
18	101.5	3.7	926	2	US-08-855-910-2	Sequence 2, Appli
19	101	3.7	721	2	US-09-583-110-5179	Sequence 5179, Ap
20	101	3.7	750	2	US-09-107-433-4321	Sequence 4321, Ap
21	101	3.7	1698	2	US-09-902-540-12813	Sequence 12813, A
22	100.5	3.7	468	2	US-09-832-129-59	Sequence 59, Appl
23	100	3.7	569	2	US-08-961-083-154	Sequence 154, App
24	100	3.7	569	2	US-09-536-784-154	Sequence 154, App
25	100	3.7	569	2	US-09-765-271-154	Sequence 154, App
26	100	3.7	569	2	US-09-765-272A-154	Sequence 154, App
27	100	3.7	658	2	US-09-769-787-17	Sequence 17, Appl
28	677	100	3.7	2	US-09-769-787-155	Sequence 155, App
29	752	98	3.6	2	US-09-252-991A-30127	Sequence 30127, A
30	511	97.5	3.6	2	US-09-266-965-127	Sequence 127, App
31	1047	97.5	3.6	2	US-09-252-991A-29125	Sequence 29125, A
32	432	96.5	3.6	2	US-09-832-129-41	Sequence 41, Appl
33	432	96.5	3.6	2	US-09-832-129-54	Sequence 54, Appli
34	438	96.5	3.6	2	US-09-167-299-5	Sequence 5, Appli
35	4545	96.5	3.6	1	US-08-804-227C-14	Sequence 14, Appl
36	4550	96.5	3.6	1	US-08-804-227C-8	Sequence 8, Appli
37	4550	96.5	3.6	1	US-08-804-198-2	Sequence 2, Appli
38	823	96	3.5	2	US-09-252-991A-24768	Sequence 24768, A
39	301	95.5	3.5	2	US-09-148-545-232	Sequence 232, App
40	301	95.5	3.5	2	US-09-621-011-232	Sequence 232, App
41	639	95.5	3.5	2	US-09-252-991A-25089	Sequence 25089, A
42	591	95	3.5	2	US-08-961-083-74	Sequence 74, Appl
43	591	95	3.5	2	US-09-536-784-74	Sequence 74, Appl
44	591	95	3.5	2	US-09-765-271-74	Sequence 74, Appl
45	591	95	3.5	2	US-09-765-272A-74	Sequence 74, Appl

ALIGNMENTS

RESULT 1
US-09-687-360-2
; Sequence 2, Application US/0968736O
; Patent No. 6617143
; GENERAL INFORMATION:
; APPLICANT: Fukuyama, Shiro
; TITLE OF INVENTION: Polypeptides Having Glucanotransferase
; TITLE OF INVENTION: Activity And Nucleic Acids Encoding Same
; FILE REFERENCE: 6012.200-US
; CURRENT APPLICATION NUMBER: US/09/687,360
; CURRENT FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: PA 1999 01501
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: PA 1999 01641
; PRIOR FILING DATE: 1999-11-15
; PRIOR APPLICATION NUMBER: 60/160,903
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: 60/166,539
; PRIOR FILING DATE: 1999-11-18
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 2
;   LENGTH: 501
;   TYPE: PRT
;   ORGANISM: Thermus rubens
US-09-687-360-2

Query Match   100.0%; Score 2714; DB 2; Length 501;
Best Local Similarity   100.0%; Pred. No. 6.5e-252;
Matches 501; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MQLQRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQ 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MQLQRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQ 60

Qy   61 SFSAFAGNPYLVDPEMLIEKGWLEQSEAPPYPTQRVDYGWLYQTRWPLLRRAFAGFRAR 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 SFSAFAGNPYLVDPEMLIEKGWLEQSEAPPYPTQRVDYGWLYQTRWPLLRRAFAGFRAR 120

Qy 121 ASAQDKTRLEAPIEAARFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAREEL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 ASAQDKTRLEAPIEAARFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAREEL 180

Qy 181 AEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 AEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADG 240

Qy 241 NPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGF 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 NPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGF 300

```
Qy   301  EAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGF  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  EAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGF  360

Qy   361  PGMKIIQFAFSGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAYLA  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  PGMKIIQFAFSGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAYLA  420

Qy   421  RYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWANRYAE  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  RYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWANRYAE  480

Qy   481  GDLEPGLAAGLRALAEASQRA  501
          |||||||||||||||||||||
Db   481  GDLEPGLAAGLRALAEASQRA  501


RESULT 2
US-09-583-110-4228
; Sequence 4228, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4228
; LENGTH: 505
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-4228

Query Match             44.1%;  Score 1196;  DB 2;  Length 505;
Best Local Similarity   48.1%;  Pred. No. 4.1e-106;
Matches  242;  Conservative   77;  Mismatches  168;  Indels   16;  Gaps    9;

Qy     4  QRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQSFS   63
          :|:|::| ::|||:| |||: |:.:: : ||:: : ||| :||||.|||||||||||||
Db     3  KRQSGVLMHISSLPGAYGIGSPGQSAYDFVDFLVRTKQRYWQILPLGATSYGDSPYQSFS   62

Qy    64  AFAGNPYLVDPEMLIEKGWLEQSEAPP---PYPTQRVDYGWLYQTRWPLLRRAFAGFRAR  120
          ||||| :.:|: :|:|:| ||: . |.    : .  ||||: :  :||||.:: .:|:.
Db    63  AFAGNTHFIDLDILVEQGLLEASDLEGVDFGSDASEVDYAKIYYARRPLLEKAVKRFFEV  122

Qy   121  ASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEW-SPELRDREPAALARAREE  179
          .. :| :  :.| ||| .|:| :|..:.|| .  .  | ||   |:||     ::  | |
Db   123  GDVKD---FEKFAQDNQSWLELFAEYNAIKEHFDNLAWTEWPDADARARKASALESYREQ  179

Qy   180  LAEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEAD  239
          |::: . : : :|:|.  .:|.||| ::::.|:|:||||:||.|||| ||||: |: .:
Db   180  LADKLVVHRVTQYFFFQQWLKLKAYANDNHIEIVGDMPIYVAEDSSDMWANPHLFKTDVN  239

Qy   240  GNPTVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRG   299
          |: . ||.| |:|| |||||||||:.|:.|::|:: |  ||:|:|:|| .:||||||||
Db   240  GKATCIAGCPPDEFSVTGQLWGNPIYDWEAMDKDGYKWWIERLRESFKIYDIVRIDHFRG  299

Qy   300  FEAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFG  359
          ||:|||:|.| .:| ||.||| || :|||||:.: .:   ||||||::|:|:| |::. |
Db   300  FESYWEIPAGSDTAAPGEWVKGPGYKLFAAVKELGELNIIAEDLGFMTDEVELRERTG    359

Qy   360  FPGMKIIQFAFSGEDNAF-LPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAY  418
          |||||::||||: ||:     |   | .| . :.|||||:|.||| |:    :||  | |
Db   360  FPGMKILQFAFNPEDESIDSPHLAPA--NSVWYTGTHDNNTVLGWYRN--EIDDA-TREY  414
```

```
Qy   419  LARYGIRCLSEYE-VAGALIBLAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWR  477
          :|:|  :  .::  : | |:. |  :. .:: | ||| ::|. :|| :||||.|| |:|.|:
Db   415  MARYTNR--KEYETVVHAMLRTVFSSSVFMAIATMQDLLELDEAARMNFPSTLGGNWSWR  472

Qy   478  YAEGDLEPGLAAGLRALAEASQR  500
          :|:   .| | ||||  |: ::|:
Db   473  MTEDQLTPAVEBGLLDLTTIYRR  495


RESULT 3
US-09-107-433-2939
; Sequence 2939, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
;   SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE        FOR DIAG[
;   THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;   STREET: 100 Beaver Street
;   CITY: Waltham
;   STATE: Massachusetts
;   COUNTRY: USA
;   ZIP: 02354
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: CD/ROM ISO9660
;   COMPUTER: <Unknown>
;   OPERATING SYSTEM: <Unknown>
;   SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/107,433
;   FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 60/ 085131
;   FILING DATE: May 12, 1998
;   APPLICATION NUMBER: 60/051553
;   FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
;   NAME: Ariniello, Pamela Deneke
;   REGISTRATION NUMBER: 40,489
;   REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (781)893-5007
;   TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 2939:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 458 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
; MOLECULE TYPE: protein
;   HYPOTHETICAL: YES
;   ORIGINAL SOURCE:
;   ORGANISM: Streptococcus pneumoniae
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (B) LOCATION 1...458
; SEQUENCE DESCRIPTION: SEQ ID NO: 2939:
US-09-107-433-2939

Query Match             39.6%;  Score 1076;  DB 2;  Length 458;
Best Local Similarity   48.4%;  Pred. No. 1.2e-94;
Matches  221;  Conservative   66;  Mismatches  154;  Indels   16;  Gaps    9;

Qy    50  GPTSYGDSPYQSFSAFAGNPYLVDPEMLIEKGWLEQSEAPP---PYPTQRVDYGWLYQTR  106
          | ||||||||||||||||| :.:|: :|:|:| ||: . |.    : .  ||||: :
Db     2  GATSYGDSPYQSFSAFAGNTHFIDLDILVEQGLLEASDLEGVDFGSDASEVDYAKIYYAR   61

Qy   107  WPLLRRAFAGFRARASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEW-SPEL  165
          :|||.:: .:|:. .. :| :  :.| ||| .|:| :|..:.|| .  .  | ||
Db    62  RPLLEKAVKRFFEVGDVKD---FEKFAQDNQSWLELFAEYMAIKEHFDNLAWTEWPDADA  118
```

```
Qy    3 IEIVGDMPIYVAEDSSDWANPHLFKTDVNGKATCIAGCPPDEFSVTGQLWGNPIYDWEA  62
Db  270 MERDNFAWCIARIRQSLKQCHLVRIDHFRGFEAXWEVPFGRPNAVEGRWVKAPGEKLFAA 329

Qy   63 MDKDGYKWWIERLRESFKIYDIVRIDHFRGFESYWEIPAGSDTAAPGEWVKGPGYKLFAA 122
Db  330 VRAQLSDAPIIAEDLGVITPEVEALRDGFGFPGMKILQFAFSGEDNAF-LPHNYPAHGNV 388

Qy  123 VKEELGLNIIAEDLGPMTDEVIELRERTGFPGMKILQFAFNPEDESIDSPHLAPA--NS  180
Db  286 LKQCHLVRIDHFRGFEAXWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLG 345

Qy  389 VVYSGTHDNDTTLGWFRTAPEAERAFMRAYLARYGIRCLSEYE-VAGALIELAFKSPAKL 447
Db  181 VMYTGTHDNNTVLGWYRN--EIDDA-TREYMARYTNR--KEYETVVHAMLRTVFSSVSFM 235

Qy  448 AIVPLQDVLGLGPEARMNFPGRLGDNWAWRYAEGDLEPGLAAGLRALAEASQR 500
Db  236 AIATMQDLLELDEAARMNFPSTLGGNWSWRMTEDQLTPAVEEGLLDLTTIYRR 288
```

RESULT 5
US-08-838-543-6
; Sequence 6, Application US/08838543
; Patent No. 594623
; GENERAL INFORMATION:
;   APPLICANT: KREBBERS, ENNO
;   APPLICANT: BROGLIE, KAREN E.
;   TITLE OF INVENTION: CORN 4-(-GLUCANOTRANSFERASE
;   NUMBER OF SEQUENCES: 6
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: E. I. DUPONT DE NEMOURS AND COMPANY
;   STREET: 1007 MARKET STREET
;   CITY: WILMINGTON
;   STATE: DELAWARE
;   COUNTRY: UNITED STATES OF AMERICA
;   ZIP: 19898
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3.50 INCH DISKETTE
;   COMPUTER: IBM PC COMPATIBLE
;   OPERATING SYSTEM: MICROSOFT WINDOWS 95
;   SOFTWARE: MICROSOFT WORD FOR WINDOWS 95 (VER. 7.0)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/838,543
;   FILING DATE:
;   CLASSIFICATION: 800
;   ATTORNEY/AGENT INFORMATION:
;   NAME: MAJARIAN, WILLIAM R.
;   REGISTRATION NUMBER: P-41,173
;   REFERENCE/DOCKET NUMBER: BB-1101
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 302-992-4926
;   TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 291 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
US-08-838-543-6

Query Match   25.6%; Score 694; DB 1; Length 291;
Best Local Similarity 50.2%; Pred. No. 2.9e-58;
Matches 136; Conservative 40; Mismatches 87; Indels 8; Gaps 3;

```
Qy  208 KGIQIIGDMPIFVAFDSSDWANPQYFYLEADGNPTVVAGVPRDYFSETGQLWGNPLYRW  267
Db    1 KGISIMGDMPIYVGYHSADVWANKKQFLLNRKGFPLIVSGVPPDAFSETGQLWGMGSPLYDW 60

Qy  268 DVMERDNFAWCIARIRQSLKQCHLVRIDHFRGFEAXWEVPFGRPNAVEGRWVKAPGEKLF 327
Db   61 KAMEKDGFSWWVRRIQRATDLFDEFRIDHFRGFAGFWAVPSEEKIAILGRWKVGPGKPLF 120

Qy  328 AAVRAQLSDAPIIAEDLGVITPEVEALRDGFGFPGMKILQFAF-SGEDNAFLPHNYPAHG 386
```

```
Qy  166 RDREPAALARARELAEEVALXEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSS 225
Db  119 RARKASALESYRRQOLADKLVYHRVTQYFFQQWLKLKAYANDNHIEIVGDMPIYVAEDSS 178

Qy  226 DVWANPQYFYLEADGNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQS 285
Db  179 DMWANPHLFKTDVNGKATCIAGCPPDEFSVTGQLWGNPIYDWEAMDKDGYKWWIERLRES 238

Qy  286 LKQCHLVRIDHFRGFEAXWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLG 345
Db  239 FKIYDIVRIDHFRGFESYWEIPAGSDTAAPGEWVKGPGYKLFAAVKEELGLNIIAEDLG 298

Qy  346 VITPEVEALRDGFGFPGMKILQFAFSGEDNAF-LPHNYPAHGNV 404
Db  299 FMTDEVIELRERTGFPGMKILQFAFNPEDESIDSPHLAPA--NSVMTGTHDNNTVLGWY 356

Qy  405 RTAPEAERAFMRAYLARYGIRCLSEYE-VAGALIELAFKSPAKLAIVPLQDVLGLGPEAR 463
Db  357 RN--EIDDA-TREYMARYTNR--KEYETVVHAMLRTVFSSVSFMAIATMQDLLELDEAAR 411

Qy  464 MNFPGRLGDNWAWRYAEGDLEPGLAAGLRALAEASQR 500
Db  412 MNFPSTLGGNWSWRMTEDQLTPAVEEGLLDLTTIYRR 448
```

RESULT 4
US-08-838-543-5
; Sequence 5, Application US/08838543
; Patent No. 594623
; GENERAL INFORMATION:
;   APPLICANT: KREBBERS, ENNO
;   APPLICANT: BROGLIE, KAREN E.
;   TITLE OF INVENTION: CORN 4-(-GLUCANOTRANSFERASE
;   NUMBER OF SEQUENCES: 6
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: E. I. DUPONT DE NEMOURS AND COMPANY
;   STREET: 1007 MARKET STREET
;   CITY: WILMINGTON
;   STATE: DELAWARE
;   COUNTRY: UNITED STATES OF AMERICA
;   ZIP: 19898
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3.50 INCH DISKETTE
;   COMPUTER: IBM PC COMPATIBLE
;   OPERATING SYSTEM: MICROSOFT WINDOWS 95
;   SOFTWARE: MICROSOFT WORD FOR WINDOWS 95 (VER. 7.0)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/838,543
;   FILING DATE:
;   CLASSIFICATION: 800
;   ATTORNEY/AGENT INFORMATION:
;   NAME: MAJARIAN, WILLIAM R.
;   REGISTRATION NUMBER: P-41,173
;   REFERENCE/DOCKET NUMBER: BB-1101
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 302-992-4926
;   TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 298 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
US-08-838-543-5

Query Match   28.6%; Score 777.5; DB 1; Length 298;
Best Local Similarity 53.2%; Pred. No. 2.9e-66;
Matches 156; Conservative 39; Mismatches 89; Indels 9; Gaps 6;

```
Qy  210 IQIIGDMPIFVAFDSSDVWANPQYFYLEADGNPTVVAGVPRDYFSETGQLWGNPLYRWDV 269
```

```
Db  121 DAILQAVGKINIIAEDLGVITEDVVQLRKSIEAPGMAVLQFAFGSDAENPHLPHNH--EQ 178
             |||  |   ||||||||        :     :||||:|       |  :||||||   ||||
Qy  387 NVVVYSGTHDNDTTLGWFRTAPEAERAFMRAYLARYGIRCLSEYEVAGALIELAFKSPAK 446
        |  |: |||||||  :|||  ::    |:  :: ::||      |::|||||||: || :|
Db  179 NQVVYTGTHDNDTIRGWNDTLPQEEKSNVLKYLSN----IEEEISRGLIEGAVSSVAR 233

Qy  447 LAIVPLQDVLGLGPEARMNFPGRLGDNWAWR 477
        :|:||:||||||| ::|||| ||::|||:||
Db  234 IAIIPMQDVLGLGSDSRMNIPATQFGNWSWR 264
```

RESULT 6
US-09-902-540-9896
; Sequence 9896, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 9896
; LENGTH: 470
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-9896

Query Match 22.3%; Score 606.5; DB 2; Length 470;
Best Local Similarity 34.4%; Pred. No. 1.5e-49;
Matches 158; Conservative 73; Mismatches 178; Indels 53; Gaps 16;

```
Qy   46 VLPLGPTSYGD-SPYQSFSAFAGNPYLVDPEMLIE----------------KGWLEQSEAPP  90
        ||||  |: || ||| :  |||| || | |   :|:                ||||||||||
Db    2 VLPLLPTAPGDSSPYAITRSAFGLNPLFIDLQRLPEFTAAGGEAALSDAQRAQLAEARSAP  61

Qy   91 PYPTQRVDYGWLYQTRWPLLRRAFAGFRARASAQDKTRLEAF--IEBAERFWLEDYALFM 147
            :  | |  :  |   || |||||||||||||||   |   |  |:  ||||||||| ||
Db  -62 -----RVRYDLVFPLKDTPAFARAFDVFEAQHWAPQSERAKTFPQWRDAQGEWIESYALFT 116

Qy  148 ALKTRFDGKPWNEWSPELRDREPAALARAREELAEEVALYEWIQWLFYLEWGQTKAYAES 207
        |:  :    |W EW  |LR R:P|||  :   ||:     : : |:| ||  :  | |  :
Db  117 AISEQEERRPWWEWPEPLRTRQPEALAQQKSRELERRVRYHAWLQWVAEQQWNEVRTQARA 176

Qy  208 KGIQIIGDMPIFVAFDSSDVWANPQYFYLEADGNPTVVAGVPRDYFSETGQLWGNPLYRW 267
        | : : |   | : | |:DVWA:P  : |  | ||    || || | |||: |  | :
Db  177 KDVLLCGDEPFIIGQDSADVWAHPDILRRDAR------LGVPPDDFSATGQDWGLPYFDL 230

Qy  268 DVMERDNFAWCIARIRQSLKQCHLVRIDHFRGFEAYW---EVPFGR---PNAVEGRWVK 320
         ::| |:FAW        :  : ||:VDH  G  : |    :  | ::   :  | |
Db  231 AAMEKDDFAWLKTRAKKAASYYDLRRVDHAVGYFRQWIRDAQTPTGRFIPPDEESHRRL- 289

Qy  321 APGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGFPGMKILQFAFSGEDNAFL-P 379
          ||| :   | :  ::| |AEDLGVI    : | | :| | |:::  :   :   :  |
Db  290 --GEKTF---RLLSEGAGIVAEDLGVIPHFVRHILAELKLPGYRVMR--WERDDNHYRDP 342

Qy  380 HNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAER-AFMRAYLARYGIRCLSEY--EVAGAL 436
        ::  |   ||  :GTHD DT   W  |  ||:: | :| :   |      |    :|  :
Db  343 RQFPAVS--LVTTGTHDTDTMAEWWEGAGHERQAAARSWPELNGVAITREFTPDVHRAM 400

Qy  437 IELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGD-NWAWR 477
        :   :  : |  : |||| || |   ||:| |:|    ||::
Db  401 LASALNANSDLCLVLPWQDV--LGTRDRINLPGSMSDSNWAYR 440
```

RESULT 7
US-09-198-452A-347

; Sequence 347, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragment
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prev
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 347
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-347

Query Match 16.6%; Score 451; DB 2; Length 526;
Best Local Similarity 26.8%; Pred. No. 1.6e-34;
Matches 131; Conservative 87; Mismatches 212; Indels 58; Gaps 15;

```
Qy   21 GIGALGREAERFLD------WLADAGARWWQVLPLGPTSYGDSPYQSFSAFAGNPYLVDP  74
        ||| |                  :| ||  ||||| |||||:|||||||||||||||||||
Db   41 GIG-------EFLDLIPLISWCCKQGFSVIQLLPLNDTGEDTSPYNSISSVALNPLFLSL  93

Qy   75 EML--------IEKGWLEQSEAPPPYPTQRVDYGWLYQTRWPLLRRAFAGFRARASAQDK 126
        |:|        :|:|||||||   |||||||||||||||||||||||||||||||||||||
Db   94 SSLPNIDTIPEVAKKLQDMHEL---CSTPSVSVTQVKEKKWAFLREYYQKC-CKSSLEGN 149

Qy  127 TRLEAFIRAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAREELAEEVAL 186
         : | |: :||:WL |   |:|:     ||| ||W      |: |   | ::  ||:
Db  150 SNFSEFLESERYWLYPYGTFRAIKHHMGEPINNWPKSLTDQE--NFPDLTKKFHDEVLF 207

Qy  187 YEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADGNPTVVA 246
        :  :|||  :   :  ||| :: : :: || ||   |  | :    :   : |  : :|
Db  208 FSYLQFLCYQQLCEVKAYADQHHVLLKGDLPILISKDSCDVWYFRDYF------SSSRSV 261

Qy  247 GVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGFEAYWEV 306
        | | ||  | ||:|  |:| :::||:||  :     | : ::  :   |||  |||  :|
Db  262 GAPDLYNSEGQNWHLPIYNFSQLAKDDYIWWKERLRYAQNFYSVYRLDHIIGFFRLWI- 320

Qy  307 PFGRPNAVEGRWVK--------APGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFG 359
                      :          |  | : | :  ::  : |||||| :  | ||||  ||
Db  321 --WDSSGRGRFIPDNPKDYIKQGTEILSTMLGASSMLP-IGEDLGIIPQDVKTTLTHLG 376

Qy  360 FPGMKILQFAFSGE-DNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAY 418
        : | :|  |||  | ::|::|   :||||  | ||||:||  |  : |:  :  : |
Db  377 ICGTRIPRWERNWESDSAFIPLKDYNPLSVTTLS-THDSDTFAQWMLNSPKEAKQFAKFL 435

Qy  419 LARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPE-------ARMNFPGRL 470
            :|: ||| : :|  :|  :  :   :   ||  |:| | |        |:||||| |
Db  436 HLPFQKTLTTETQI--DILKLSHSASIFHINLFNDYLALCPDLVSKNLQRERINTPGTI 493

Qy  471 G-DNWAWR 477
        | :: |:|
Db  494 SKKNWSYR 501
```

RESULT 8
US-09-438-185A-328
; Sequence 328, Application US/09438185A
; Patent No. 6822071
; GENERAL INFORMATION:
; APPLICANT: Stephens, Richard Wayne
; APPLICANT: Mitchell, Wayne
; APPLICANT: Kalman, Sue
; APPLICANT: Davis, Ronald
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Chlamydia Pneumoniae Genome Sequence
; FILE REFERENCE: 018941-0041US
; CURRENT APPLICATION NUMBER: US/09/438,185A
; CURRENT FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: US 60/108,279

```
; PRIOR FILING DATE: 1998-11-12
; PRIOR APPLICATION NUMBER: US 60/128,606
; PRIOR FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 1074
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 328
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; OTHER INFORMATION: CPn0326
US-09-438-185A-328

Query Match       16.6%;  Score 451; DB 2; Length 536;
Best Local Similarity   26.8%;  Pred. No. 1.6e-34;
Matches  131;  Conservative   87;  Mismatches  212;  Indels   58;  Gaps   15;

Qy   21 GIGALGREAERFLD------WLADAGARWWQVLPLGPTSYGDSPYQSFSAFAGNPYLVDP  74
Db   51 GIG-------EFLDLIPLISWCQKQGFSVIQLLPLNDTGEDTSPYNSISSVALNPLFLSL 103

Qy   75 EML----------IEKGWLEQSEAPPPYPTQRVDYGWLYQTRWPLLRRAFAGFRARASAQDK 126
Db  104 SSLPNIDTIPEVAKKLQDMHEL---CSTPSVSYTQVKEKKWAFLREYYQKC-CKSSLEGN 159

Qy  127 TRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAREELABEVAL 186
Db  160 SNFSEFIESERYWLYPYGTFRAIKHHMGEPINNWPKSLTDQE--NFPDLTKKFHDEVLF 217

Qy  187 YEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADGNPTVVA 246
Db  218 FSYLQFLCYQQLCEVKAYADQHHVLLKGDLPILISKDSCDVWYFRDYF-------SSRSV 271

Qy  247 GVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGFEAYWEV 306
Db  272 GAPPDLYNSEGQNWHLPIYNFSQLAKDDYIWWKERLRYAQNFYSVRLDHIIGFFRLWI- 330

Qy  307 PFGRPNAVEGRWVK-------APGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFG 359
Db  331 ---WDSSGRGRFIPDNPKDYIKQGTEILSTMLGASSMLP-IGEDLGIIPQDVKTTLTHLG 386

Qy  360 FPGMKILQFAFSGE-DNAFLPHNYPAHGNVVYSGTHDNDTTLGWFRTAPEAERAFMRAY 418
Db  387 ICGTRIPRWERNWESDSAFIPLKDYNPLSVTTLS-THDSDTFAQWWLNSPKEAKQFAKFL 445

Qy  419 LARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPE-------ARMNFPGRL 470
Db  446 HLPFQKTLTTETQI--DILKLSHESASIFHINLFNDYLALCPDLVSKNLQRERINTPGTI 503

Qy  471 G-DNWAWR 477
Db  504 SKKNWSYR 511


RESULT 9
US-09-252-991A-24629
; Sequence 24629, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24629
; LENGTH: 681
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24629

Query Match       11.0%;  Score 299; DB 2; Length 681;
Best Local Similarity   24.9%;  Pred. No. 9.1e-20;
Matches  135;  Conservative   69;  Mismatches  252;  Indels   69;  Gaps   21;

Qy   17 PGRWGIGALGREAERFLDWLADAGARWWQVLPLGP------TSYGDSPYQSFSAFAGNP  69
Db  160 FGDGGLG---DSAALEDLLRSAARHGADALAISPLHALAEANGHAYSPYSPSSRLFFNV 215

Qy   70 YLVDPEMLIEKGWLEQ-------SEAPPPYPTQRVDYGWLYQTRWPLLRRAFAGFRARA 121
Db  216 LHAAPATILGAAAVEQAIRRAGLAAEMARLESLELIDWTAAADLRMRLLRQLHRDFTERA 275

Qy  122 SAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKP-WNEW-SPELRDREPAALARAREE 179
Db  276 SPLRHDLAEFRREAGEALLH-HCRFETLQAHLGAGPDWRRWPEPLRRPGEPAVAAFCADH 334

Qy  180 LABEVALYEWIQWLFY--LEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLE 237
Db  335 -ABEVDFRAFGQWLTQRCLQHAQRQARREAGMAIGLVADLAVGADGGGSQAWSRQEELLAE 393

Qy  238 ADGNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHF 297
Db  394 ------VNVGAPPDILNQSGQDWGVSAFNPEGLRRHGYRAFREMLRANLAWPGGLRIDHV 447

Qy  298 RGFEAYWEVPFGRPNAVEGRWVKAPGEKL--FAAVRAQLSDAPIIAEDLGVITPEVEALR 355
Db  448 MGLQRLWLIPRGQPPHA-GAYLRYPQRELLRLLALEASRASALVIGEDLGTVP---EGLR 503

Qy  356 DGFG--FPGMKILQFAFSGEDNAFl-PHNYPAHGNVVYSGTHDNDTTLGWFRTAPEAE 411
Db  504 EELARRQVLGTRVLLFERRGE--RFVPPAQWPA--DAMATTSTHDLFSLSGWWRGQDIHW 559

Qy  412 RAFMRAYLARYGIR----CLSEYEVAG------------ALIELAFK 442
Db  560 R------GRAGHRSAEECAADLELRABERRALAASLEPPVDPDPAAEVPLDACIGYVGA 612

Qy  443 SPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWRYAEGDLE----PGLAAGLRALAEAS 498
Db  613 TPAPLVLLPLEDALGSLEQPNLPGPGDAHPNWRRRWPENAAQMLGTPQVDRRLR-LLERS 671

Qy  499 QR 500
Db  672 RR 673


RESULT 10
US-09-489-039A-11073
; Sequence 11073, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 11073
; LENGTH: 743
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-11073

Query Match       10.2%;  Score 278; DB 2; Length 743;
Best Local Similarity   23.3%;  Pred. No. 1.1e-17;
Matches  121;  Conservative   72;  Mismatches  247;  Indels   80;  Gaps   17;

Qy   20 WGIGALGREAERFLDWLADAGARW------WQVLPLGPTSYGDSPYQSFSAFAGNPYLVD  73
```

```
        MOLECULE TYPE:  protein
  US-08-838-543-2

  Query Match                 10.1%;    Score 273.5;  DB 1;  Length 496;
  Best Local Similarity       22.1%;    Pred. No. 1.6e-17;
  Matches  100;  Conservative   38;  Mismatches 139;  Indels 175;  Gaps  14;

Qy   188  EWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADGNPTVVAG  247
                :::!::: :::!!                    ::!:!:::!!::            ::  :
Db     5  DYVQYHLYIQLSEAATYARKKNVILKGDLPIGVDRNSVDTWVYPTLFRM------NTATG   58

Qy   248  VPRDYFSETGQLWGNPLYRWDVMERDNPAWCIARIRQSLKQCHLVRIDHFRGFEAYWEVP  307
              ::!!  ::: !!:!!!::::: :: !::::::::::::!!:::::::!!!:!!:!!:!
Db    59  APPDYFDKNGQNWGFPTYNWEEMSKDNYGWRARLTQMAKYFTAYRIDHILGFFRIWELP  118

Qy   308  ------FG--RPN------AVEGRW------                             318
                !!::                :::
Db   119  DHAATGLVGKFRPSIPLSQEELLSEGLWDFNRMSQPYIRQEILEEKFGSFWTVIAANFLN  178

Qy   319  ------VKAPGEK------LFAAV------                              330
                 !!::                 :
Db   179  EYQKQCYEFKEDCNTEKKIIVKIKTYSAEKSLWVEKEDNIRRGLFDLLQNIVLIRDPEDST 238

Qy   331  ---RAQLSDA----------PII---                                  340
                 !!:                ::
Db   239  KFYPRFNLEDTSSFRDLDEHSKNILRRLYYNYFVRQENLWRQNALKTLPVLLNSSDMLA  298

Qy   341  -AEDLGVITPEVEALRDGFGPPGMKILQFAFSGEDNAFLPHNYPAHGNVVVSGTHDNDT  399
               !!!!                      ::!:!!::::!::!:::::::::  :::::::
Db   299  CGEDLGLIPACVHPVMQELGLIGLRIQRMPSEPNLEFGIPSQYSY--MTVCAPSCHDCST 356

Qy   400  TLGWFRTAPAERAFMRAYLARYG------IRCLSEYEVAGALIELAFKSPAKLAIVPLQ  453
                :::                        !::::::::!:::::::::::::::::::::
Db   357  LRAWWE--EDEGRRSRFYKTVVGSDEEPPSRCTP--EVVHFIVQQHFDAPSMWAIFPLQ  411

Qy   454  DVLGL------------GPEARMNFPGRLGDNWAWR                        477
            :!:               !:!::::::::::::::
Db   412  DLLALKDKYTTRPAPEETINDPTNPKHYWRFR                            443
```

```
Db   206  WGIGDFG-DLRAMLPEIARRGGSFIGLNPIHALYPANPES--ASPYSPSSRRWLNVIYID  262
            !!!!    :   !      !   !   :!    ::!        :!         ::!
Qy    74  PEMLIEKGWLEQSEAPPPYP--------TQRVDYGWLYQTRWPLLRRAFAGFRARAS    122
            !!!                         :!!!  !!!  !::!:!::::::::
Db   263  VNAVEDFQRSEEAQAWWQSPATQQALQAARETDDVDYTAVTTLKMTALRMAWKQFSRRED  322

Qy   123  AQDKTRLEAFIEAE--RFWLEDYALFMALKTRFDGKPWN--EWSPELRDREPAALARAR  177
            !:::              ::  :  :                  :
Db   323  EQ-MTAFREFVLREGESLYWQAAFDALHAWQVQQDPLRWGPAWPKAFQDIDSPEVKAFC  381

Qy   178  EELAEEVALYEWIQWLFYLEWGQTKAYAESK-----GIQIIGDMPIFVAFDSSDVWANPQ  232
            : :: !::!:! :::!!!            :
Db   382  VEHEDDVSFYLWLQW---LAWSQFAACWETSQRDGMPIGLYRDLAVGVAEGGSETWCDRE  438

Qy   233  YFYLEADGNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLV  292
            ::            !!  ::: !!:!!!::::: ::!::::::::::::::!!:::::::
Db   439  LYCLKAS------VGAPPDILGPLGQNWGLPPMDPHIIARAYEPFIDLLRANMQNCGAL  492

Qy   293  RIDHFRGFEAYWEVPFGRPNAVEGRWVKAPGEKLFA--AVRAQLSDAPIIAEDLGVITPE  350
            !!!:!!:!!:!!:!       !!::       :       !!:         !!!!
Db   493  RIDHVMSVLRLWWIFYGE-TADHGAYVQYPVDDLLSLLALESQRHRCMVIGEDLGTVPVE  551

Qy   351  VEALRDGFGPPGMKILQFAFSGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEA  410
                         :!:!!::::!::!:::::::::  ::::::::  :::
Db   552  IVSKLRKSGVYSYKVLYFESDAEKTPRAPALYPEQSMAV-ATTHDLPTLRGYWESGDLT  609

Qy   411  ERAFMRAY------------LARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQD--  454
            ::                  !::::::::!:::::::::::::::::::::::
Db   610  LGKALGLYPDEVLRGLYQDRELAKQGL-LDALHKYGCLPKRAGHKASLMSMTGILNRG  667

Qy   455  ------------VLGLGPE------ARMNFPGRLGDNWAWR                   477
                        !:!::                ::::::::::::
Db   668  MQRYIADSNSALLGLQPEDWLEMATPVNIPGTSTEYPNWR                    707
```

```
;      SEQUENCE CHARACTERISTICS:
;        LENGTH: 323 amino acids
;        TYPE: amino acid
;        STRANDEDNESS: single
;        TOPOLOGY: linear
;      MOLECULE TYPE: peptide
US-08-838-543-4

Query Match             7.3%;  Score 197;  DB 1;  Length 323;
Best Local Similarity  29.6%;  Pred. No. 1.9e-10;
Matches  56;  Conservative  33;  Mismatches  84;  Indels  16;  Gaps   6;

Qy  212 IIGDMPIFVAFDSSDVWANPQYFYLEADGNPTVVAGVPRDYFSETGQLWGNPLYRWDVME 271
            ||:||||.:|||||:  :  :|||  |||||||||||||| |||||||  :||
Db    7 IYGDLAVNSSRGSADVWSDPDLYCVNAS------IGAPPDLGPVGQNWNLPPYNPTVLK  60

Qy  272 RDNFAWCIARIRQSLKQCHLVRIDHFRGFEAYWEVPFGRPNAVEGRWVKAPGEKLFA--A 329
            ||  ||  ::| |  ||||||||||||||||||||||||||||||||||||  :||
Db   61 ARGFAPFIDMLCANMQYFGVLRIDHVMGLFRLMWIPKGK-TAADGAYHYPFDELMAILA 119

Qy  330 VRAQLSDAPIIAEDLGVITPEVEALRDGFGFPGMKILQFAFSGEDNAFLPH--NYPAHGN 387
            :|:||.| :|||||  ||| :|||||||:|||||||||||||||| ||   :||
Db  120 IESVRNECLIIGEDLGTVPDEVRWKLNEFQIFSYFVLYFA---QRNGEFPRISDYPRNAY 176

Qy  388 VVVYSGTHD 396
            :||||
Db  177 ATI--GTHD 183


RESULT 13
US-08-838-543-3
;  Sequence 3, Application US/08838543
;  Patent No. 5994623
;  GENERAL INFORMATION:
;      APPLICANT: KREBBERS, ENNO
;      APPLICANT: BROGLIE, KAREN E.
;      TITLE OF INVENTION: CORN 4-(-GLUCANOTRANSFERASE
;      NUMBER OF SEQUENCES: 6
;      CORRESPONDENCE ADDRESS:
;        ADDRESSEE: E. I. DUPONT DE NEMOURS AND COMPANY
;        STREET: 1007 MARKET STREET
;        CITY: WILMINGTON
;        STATE: DELAWARE
;        COUNTRY: UNITED STATES OF AMERICA
;        ZIP: 19898
;      COMPUTER READABLE FORM:
;        MEDIUM TYPE: 3.50 INCH DISKETTE
;        COMPUTER: IBM PC COMPATIBLE
;        OPERATING SYSTEM: MICROSOFT WINDOWS 95
;        SOFTWARE: MICROSOFT WORD FOR WINDOWS 95 (VER. 7.0)
;      CURRENT APPLICATION DATA:
;        APPLICATION NUMBER: US/08/838,543
;        FILING DATE:
;        CLASSIFICATION: 800
;      ATTORNEY/AGENT INFORMATION:
;        NAME: MAJARIAN, WILLIAM R.
;        REGISTRATION NUMBER: P-41,173
;        REFERENCE/DOCKET NUMBER: BB-1101
;      TELECOMMUNICATION INFORMATION:
;        TELEPHONE: 302-992-4926
;        TELEFAX: 302-773-0164
;  INFORMATION FOR SEQ ID NO: 3:
;      SEQUENCE CHARACTERISTICS:
;        LENGTH: 330 amino acids
;        TYPE: amino acid
;        STRANDEDNESS: single
;        TOPOLOGY: linear
;      MOLECULE TYPE: peptide
US-08-838-543-3

Query Match             6.9%;  Score 187;  DB 1;  Length 330;
Best Local Similarity  23.1%;  Pred. No. 1.8e-09;
Matches  70;  Conservative  43;  Mismatches  144;  Indels  46;  Gaps   8;
```

```
Qy  210 IQIIGDMPIFVAFDSSDVWANPQYFYLEADGNPTVVAGVPRDYFSETGQLWGNPLYRWDV 269
            :  |:||| |  ::|||||||||||||||||||||||||||||||||||
Db    5 IGLYRDLAVGVTGGAETWCDRELYCLKAS-----VGAPPDIIGPLGGNWGLPPMDPHI  58

Qy  270 MERDNFAWCIARIRQSLKQCHLVRIDHFRGFEAYWEVPFGRPNAVEGRWVKAPGEKLFA- 328
            :  |:|  |  :||  | :||||||||||||||||||||||||||||||||||||
Db   59 ITARAYEPFIELLRANMQNCGALRIDHVMSMLRLWWIPY-RETADQGAYVHYPVDDLLSI 117

Qy  329 -AVRAQLSDAPIIAEDLGVITPEVEALRDGFGFPGMKILQFAPSGEDNAFLPHNYPAHGN 387
            : :: ||||.| ||||:  :|||||||:|||||||||||||||| :||:
Db  118 LALESKRHRCMVIGEDLGTVPVEIVGKLRSSGVYSYKVLYFENDHEKTFRAPKAYPEQSM 177

Qy  388 VVVYSGTHDNDT------------------TLGWFRTAPEAERAFMRAYLARYGIRCLSEYEVA 433
            :|||||||||:                   ||||||||||||||||||||||||||||
Db  178 AV--AATHDLPTLRGYWECGDLTLGKTLGLYPDEVVLRGLYQDRELAKQGL--LDALHKY 233

Qy  434 GALIELAFKSPAKLAIVPLQD-----------VLGLGPE---ARMNFPGRLGDNW 474
            :|:|||||||||||||  :|                         ||:|||||||
Db  234 GCLPKGRAGHKASLMSMTPTLNRGLQRYIADSNSALLGLQPEDWLDMAEPVNIPGTSYQYK 293

Qy  475 AWR 477
            :||
Db  294 NWR 296


RESULT 14
US-09-252-991A-25772
;  Sequence 25772, Application US/09252991A
;  Patent No. 6551795
;  GENERAL INFORMATION:
;      APPLICANT: Marc J. Rubenfield et al.
;      TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;      TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;      FILE REFERENCE: 107196.136
;      CURRENT APPLICATION NUMBER: US/09/252,991A
;      CURRENT FILING DATE: 1999-02-18
;      PRIOR APPLICATION NUMBER: US 60/074,788
;      PRIOR FILING DATE: 1998-02-18
;      PRIOR APPLICATION NUMBER: US 60/094,190
;      PRIOR FILING DATE: 1998-07-27
;      NUMBER OF SEQ ID NOS: 33142
;  SEQ ID NO 25772
;      LENGTH: 1332
;      TYPE: PRT
;      ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25772

Query Match             4.5%;  Score 123;  DB 2;  Length 1332;
Best Local Similarity  19.6%;  Pred. No. 0.02;
Matches  103;  Conservative  52;  Mismatches  155;  Indels  216;  Gaps  25;

Qy   17 PGRWGIGALGREAERF------LDWLADAGARWWQVLPLGPTSYGDSPYQSFSAFAGN  68
            ||:|: : :|| |||:           :||: |||||||||||:||:
Db   54 PGRCGAGAMS-GATLFMLVLLIVLLALLLGALGWWWR------TRGTEIRSF-------  99

Qy   69 PYLVDPEMLIEKGWLEQSEAPPPYPTQRVDYGWLYQ--------TRW--PLLRRAF 114
            ||: ||| :||   |: ||||||||||||||||                ||||||
Db  100 -YAAVRQMEREQGWQGRYESP---------------WLLMLGNETEGEQLCATWRLLPVARPAW 147

Qy  115 AGFRARASAQDKTRL---EAFIEAERF-----WLEDYALFMALKTR--FDGKPWNEWSP 163
            |||:|||:  |::|    |||:|||||      |||||||||||||  |||||||||
Db  148 FG-RWNSDGEGAILLVPESVFLPDEGLRRQSGAWLRLRLFLRLRGRALDGVVWNIPLA 206

Qy  164 ELRDREPAA---LARARE--ELAEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPI 218
            ||:||:|:    |||||    ::|:|||:||||||||||||||||||||||||||||||||
Db  207 RLQDGEQAANLGLAARRRYVELTQRLGL---------------------SLPV 238

Qy  219 FVAFDSSDVWANPQYFYLEADGNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWC 278
            ||||||||||||||||||||||||||||||||||||||||||||||||||  :::
Db  239 YVVITGLE------------DLPGFQELLAALPEE----------------- 261

Qy  279 IARIRQSLKQCHLVRIDHFRGFEAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAP 338
```

```
Db    262  ||||
           -ARER-------------ALGWSSPFAAEAAWQSRWCEQALEITATLTESIVELG  303
                             ||  |  : ||  : || :||:| |:|| :

Qy    339  I----IAEDLGVITPEVEALRDGFGPPGMKILQFAFSGEDNAFLPHNYPAHGNVVVYSGT  394
                : |:|  :|:||
Db    304  TLRGQVDNELYCLPPRLESLRGSL----QALLEPVFQGN--------ARGEAPRFRGL  349

Qy    395  HDNDTTLGWFRTAPEAERAFMRAYLARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQD  454
           :  |  ||||      :      ::|
Db    350  Y---------------------LSGSEAAGAAAD--------DVLPAVD  369

Qy    455  VLGLGPEARMNFPGRLGDNWAWRYAEGDLEPGLAAGLRALAEASQR  500
           : |:| :|  ||||
Db    370  ----APRRRSSFASQL---WARRIL--AEEGLAQAVPRILQLRQR  405


RESULT 15
US-09-583-110-5034
; Sequence 5034, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 5034
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-5034

Query Match        3.9%;  Score 105.5;  DB 2;  Length 508;
Best Local Similarity  21.8%;  Pred. No. 0.23;
Matches  73;  Conservative  51;  Mismatches 106;  Indels 105;  Gaps  22;

Qy     51  PTSYGDSPYQSFSAFAGNPYLVDPEMLIEKGWLEQSEAPPYPTQRVDYGWLYQTRWPLL  110
           |  |  :||: ||   | |   |:|: :  : :||  :: ::|| :|:| :|||:
Db     95  PLFTSDGPWRA-TLKAGT--LIEEDLFVTGNF--GSKAPYNF-SQRQEFFDEHGKKWPLM  148

Qy    111  RRAFAGFRARASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPW-NEWSPELRDRE  169
                                        |||:    |
Db    149  --------------------CMEFW-----------------DG--WFNRWKEPIITRD  168

Qy    170  PAALARAREELAEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIG--DMPIFVAFDSSDV  227
           |||   : :::                    |:   :        |   : :|
Db    169  PKELADAVREVLEQGSIN-----LYMFHGGTN-FGFMNGCSARGTLDLPQVTSYDYDAL  221

Qy    228  WANPQYFYLEADGNPT----VVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIR  283
              |: || |:  ||    : :|    | ||:  |  : ||| :
Db    222  -------LDEKRNPTAKYLAVKKMATHFSEYPQL--EPLYK-ESMELDAIP-LVEKVS  269

Qy    284  ---------QSLKQCHLVRIDHFRGP------EAYWE-------VPFGRPNA---VEG  316
                    |:  ||     || :|:   ||: :
Db    270  LFETLDSLSSPVESLYPQKMEELGQSYGYLLYRTETNWDAEEERLRIIDGRDRAQLVVDG  329

Qy    317  RWVKAP----GEKLF-AAVRAQLSDAPIIAEDLG  345
            :|||     ||| :| | :::
Db    330  QWVKTQYQTEIGEDIFYQGKKKGLSRLDLIIENMG  364
```

Search completed: January 10, 2006, 14:14:26
Job time : 49 secs

```
          GenCore version 5.1.6
     Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2006, 14:12:49 ; Search time 115 Seconds
                                       (without alignments)
                                    1820.283 Million cell updates/sec

Title:         US-10-618-976-2
Perfect score: 2714
Sequence:      1 MQLQRAFGILLHPTSFPGRW........DLEPGLAAGLRALARASQRA 501

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters:      1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    Published_Applications_AA_Main:*
         1:   /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
         2:   /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
         3:   /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
         4:   /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
         5:   /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
         6:   /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2714	100.0	501	5	US-10-618-976-2	Sequence 2, Appli
2	1194	44.0	505	5	US-10-472-928-4402	Sequence 4402, Ap
3	1076	39.6	458	5	US-10-617-320-2939	Sequence 2939, Ap
4	989	36.4	567	4	US-10-424-599-170978	Sequence 170978,
5	894.5	33.0	486	4	US-10-425-114-57203	Sequence 57203, A
6	777.5	28.6	298	4	US-10-336-597-5	Sequence 5, Appli
7	767.5	28.3	486	4	US-10-425-115-325328	Sequence 325328,
8	694	25.6	291	4	US-10-336-597-6	Sequence 6, Appli
9	477	17.6	243	4	US-10-437-963-138565	Sequence 138565,
10	451	16.6	526	3	US-09-841-132-518	Sequence 518, App
11	451	16.6	526	4	US-10-289-762-347	Sequence 347, App
12	451	16.6	526	4	US-10-872-155-518	Sequence 518, App
13	430.5	15.9	527	3	US-09-841-132-588	Sequence 588, App
14	430.5	15.9	527	5	US-10-872-155-588	Sequence 588, App
15	412.5	15.2	179	4	US-10-767-701-51019	Sequence 51019, A
16	383	14.1	683	4	US-10-081-872-188	Sequence 188, App
17	383	14.1	683	4	US-10-385-305-188	Sequence 188, App
18	373	13.7	740	4	US-10-156-761-12925	Sequence 12925, A
19	342.5	12.6	187	4	US-10-424-599-170977	Sequence 170977,
20	338	12.5	946	4	US-10-437-963-112596	Sequence 112596,
21	333	12.3	183	4	US-10-336-597-8	Sequence 8, Appli
22	323.5	11.9	204	4	US-10-425-114-55388	Sequence 55388, A
23	313.5	11.6	684	4	US-10-389-647-478	Sequence 478, App
24	273.5	10.1	496	4	US-10-336-597-2	Sequence 2, Appli
25	272	10.0	153	4	US-10-091-007-116	Sequence 116, Appli
26	270	9.9	336	4	US-10-425-115-357932	Sequence 357932,
27	238	8.8	204	4	US-10-767-701-40716	Sequence 40716, A
28	237.5	8.8	496	4	US-10-424-599-155954	Sequence 155954,
29	207	7.6	98	4	US-10-425-115-234981	Sequence 234981,
30	207	7.6	102	4	US-10-336-597-16	Sequence 16, Appl
31	207	7.6	700	4	US-10-781-014-412	Sequence 412, App
32	207	7.6	706	3	US-09-738-626-6019	Sequence 6019, Ap
33	207	7.3	706	4	US-10-781-014-410	Sequence 410, App
34	197	7.3	323	4	US-10-336-597-4	Sequence 4, Appli
35	187	6.9	330	4	US-10-336-597-3	Sequence 3, Appli
36	146	5.4	185	4	US-10-424-599-223106	Sequence 223106,
37	136	5.0	123	4	US-10-425-115-357934	Sequence 357934,
38	112	4.1	678	3	US-09-738-626-6025	Sequence 6025, Ap
39	111	4.1	595	5	US-10-472-928-4832	Sequence 4832, Ap
40	109	4.0	402	4	US-10-437-963-191559	Sequence 191559,
41	107.5	4.0	652	4	US-10-369-493-18004	Sequence 18004, A
42	105.5	3.9	284	4	US-10-425-115-367937	Sequence 367937,
43	105.5	3.9	594	5	US-10-617-320-4327	Sequence 4327, Ap
44	105	3.9	86	4	US-10-767-701-39064	Sequence 39064, A
45	105	3.9	185	4	US-10-767-701-55139	Sequence 55139, A

ALIGNMENTS

```
RESULT 1
US-10-618-976-2
; Sequence 2, Application US/10618976
; Publication No. US20050074769A1
; GENERAL INFORMATION:
; APPLICANT: Fukuyama, Shiro
; TITLE OF INVENTION: Polypeptides Having Glucanotransferase
; TITLE OF INVENTION: Activity And Nucleic Acids Encoding Same
; FILE REFERENCE: 6012 200-US
; CURRENT APPLICATION NUMBER: US/10/618,976
; CURRENT FILING DATE: 2003-07-14
; PRIOR APPLICATION NUMBER: US/09/687,360
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: PA 1999 01501
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: PA 1999 01641
; PRIOR FILING DATE: 1999-11-15
; PRIOR APPLICATION NUMBER: 60/160,903
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: 60/166,539
; PRIOR FILING DATE: 1999-11-18
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Thermus rubens
US-10-618-976-2

Query Match          100.0%;    Score 2714;  DB 5;  Length 501;
Best Local Similarity 100.0%;   Pred. No. 1e-245;
Matches  501;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1 MQLQRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQ 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MQLQRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQ 60

Qy   61 SFSAFAGNPYLVDPEMLIEKGWLEQSEAPPYPTQRVDYGWLYQTRWPLLRRAFAGFRAR 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 SFSAFAGNPYLVDPEMLIEKGWLEQSEAPPYPTQRVDYGWLYQTRWPLLRRAFAGFRAR 120

Qy  121 ASAQDKTRLEAFIEAERFWLEDYALFMALKTRPDGKPWNEWSPELRDREPAALARAREEL 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 ASAQDKTRLEAFIEAERFWLEDYALFMALKTRPDGKPWNEWSPELRDREPAALARAREEL 180

Qy  181 AEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADG 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 AEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADG 240

Qy  241 NPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGF 300
```

```
Db   360  FPGMKILQFAFSGEDNAF-LPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAY 418
          ||||||||||||:  : .|:.|||:.||||:  :||:||::|||||||| : :  ||||
Db   360  FPGMKILQFAFNPEDESIDSPHLAPA--NSVMYTGTHDNNTVLGWYRN--EIDDA-TREY 414

Qy   419  LARYGIRCLSEYE-VAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWR 477
          : ||   | :|  .:|  :  :. .:  | :|||||||:.:||: ||||||||:|||
Db   415  MARYFNR--KEYETVVHAMLRTVFSSVSFMAIATMQDLLELDEAARMNFPSTLGGNWSWR 472

Qy   478  YAEGDLEPGLAAGLRALAEASQR 500
          ||||||||||||||||||||||||
Db   473  MTEDQLTPAVEEGLLDLTTIYRR 495
```

RESULT 3
US-10-617-320-2939
; Sequence 2939, Application US/10617320
; Publication No. US2005013640A1
; GENERAL INFORMATION:
;   APPLICANT: Lynn A Doucette-Stamm   and David Bush
;   TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
;   SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE   FOR DIAGN
;   THERAPEUTICS
;   NUMBER OF SEQUENCES: 5206
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;   STREET: 100 Beaver Street
;   CITY: Waltham
;   STATE: Massachusetts
;   COUNTRY: USA
;   ZIP: 02354
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: CD/ROM ISO9660.
;   COMPUTER: <Unknown>
;   OPERATING SYSTEM: <Unknown>
;   SOFTWARE: <Unknown>
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/617,320
;   FILING DATE: 10-Jul-2003
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/09/107,433
;   FILING DATE: 30-Jun-1998
;   APPLICATION NUMBER: 60/ 085131
;   FILING DATE: May 12, 1998
;   APPLICATION NUMBER: 60/051553
;   FILING DATE: July 2, 1997
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Ariniello, Pamela Deneke
;   REGISTRATION NUMBER: 40,489
;   REFERENCE/DOCKET NUMBER: GTC-011
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (781)893-5007
;   TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 2939:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 458 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: YES
;   ORIGINAL SOURCE:
;   ORGANISM: Streptococcus pneumoniae
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (B) LOCATION 1...458
;   SEQUENCE DESCRIPTION: SEQ ID NO: 2939:
US-10-617-320-2939

Query Match   39.6%;   Score 1076; DB 5;   Length 458;
Best Local Similarity   48.4%;   Pred. No. 6.1e-92;
Matches 221; Conservative 66; Mismatches 154; Indels 16; Gaps   9;

```
Qy    50  GPTSYGDSPYQSFSAFAGNPYLVDPEMLIEKGWLEQSEAPP---PYPTQRVYDYGWLYQTR 106
          ||||||||||||||  :.:|:.:.:.|| ||  |||: :  |
```

RESULT 2
US-10-472-928-4402
; Sequence 4402, Application US/10472928
; Publication No. US2005002083A1
; GENERAL INFORMATION:
;   APPLICANT: CHIRON SpA
;   APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
;   TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
;   FILE REFERENCE: P02696WO
;   CURRENT APPLICATION NUMBER: US/10/472,928
;   CURRENT FILING DATE: 2003-09-26
;   PRIOR APPLICATION NUMBER: GB-0107658.7
;   PRIOR FILING DATE: 2001-03-27
;   NUMBER OF SEQ ID NOS: 4979
;   SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 4402
;   LENGTH: 505
;   TYPE: PRT
;   ORGANISM: Streptococcus pneumoniae
;   FEATURE:
;   OTHER INFORMATION: 4-alpha-glucanotransferase (malQ)
;   OTHER INFORMATION: Cellular location: cytoplasm
;   OTHER INFORMATION: Similar to strain R6 sequence 15903958 (0.E+01)
US-10-472-928-4402

Query Match   44.0%;   Score 1194; DB 5;   Length 505;
Best Local Similarity   48.1%;   Pred. No. 5.8e-103;
Matches 242; Conservative 77; Mismatches 168; Indels 16; Gaps   9;

```
Qy     4  QRAFGILLHPTSFPGRWGIGALGREABRFLDWLADAGARWWQVLPLGPTSYGDSPYQSFS  63
          :  | : .|| :: |||||.::|.:. :|| ||  |.||| ||||||||||||||||
Db     3  KRQSGVLMHISSLPGAYGIGSFGQSAYDFVDFLVRTKQRYWQILPLGATSYGDSPYQSFS  62

Qy    64  AFAGNPYLVDPEMLIEKGWLEQSEAPP---PYPTQRVYDYGWLYQTRWPLLRRAFAGFRAR 120
          |||||   ::|: ||||  :  ||    |||  | : ::.  | .||||||::|||||:
Db    63  AFAGNTHFIDLDILVEGLLEASDLEGVDFGGSDASEVDYAKIYYARRPLLEKAVKRFFEV 122

Qy   121  ASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEW-SPELRDREPAALARAREE 179
          ||| ::|.| ::: :.:  . |.|  :| :||   | .: || |:   | :.||.:
Db   123  GDVKD---FEKFAQDNQSWLELFAEYMAIKEYFDNLAWTEWPDADARARKASALESYREQ 179

Qy   180  LAEEVALYEWIOWLFYLEWGQTKAYABSKGIQIIGDMPIFVAFDSSDVWANPQYFYLEAD 239
          |:: : :::|  |. ||.:: |  . | | :::.. |||.|||||.|.|||||::| :
Db   180  LADKLVTHRVTQYFFQQWLKLKAYANDNHIEIVGDMPIYVAEDSSDMWANPHLFKTDVN 239

Qy   240  GNPTTVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRG 299
          ||| ||||||| :|||||||||||||||||||||||||||||||||||||||||||| |
Db   240  GKATCIAGCPPDEFSVTGQLWGNPIYDWEAMDKDGYKWWIERLRESFKIYDIVRIDHFRG 299

Qy   300  FEAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFG 359
          | ||| ||||||||||||||| |||||||:.| || ||||||||||||||||||||||
Db   300  FESYWEIPAGSDTAAPGEWVKGPGYKLFAAVKEELGELNIIAEDLGFMTDEVIELRERTG 359
```

```
Db   241  NPTTVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGF 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   301  EAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGF 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  EAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGF 360

Qy   361  PGMKILQFAFSGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAYLA 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  PGMKILQFAFSGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAYLA 420

Qy   421  RYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWRYAE 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  RYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWRYAE 480

Qy   481  GDLEPGLAAGLRALAEASQRA 501
          |||||||||||||||||||||
Db   481  GDLEPGLAAGLRALAEASQRA 501
```

```
Db     2  GATSYGDSPYQSFSAFAGNTHFIDLDILVEQGLLEASDLEGVDFGSDASEVDYAKIYYAR  61

Qy   107  WPLLRRAFAGFRARASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEW-SPEL 165
          |||.|.||::  .|:.||.:  .::|.|| ||:||  ::|.||:||:.|  :||||
Db    62  RPLLEKAVKGRFFEVGDVKD---FEKFAQDNQSWLELFAEYMAIKEHFDNLAWTEWPDADA 118

Qy   166  RDREPAALARAREELAEEVALYEWIQWLFYLLEWGQTKAYAESKGIQIIGDMPIFVAFDSS 225
          |:.|.||| ||::||:|||:::||.|||||||||||||||||||||||||||||||||||
Db   119  RARKASALESVREQLADKLVTHRVTQYFFQQWLKLKAYANDNHIEIVGDMPIYVAEDSS  178

Qy   226  DVWANPQYFYLEADGNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQS 285
          ||||||||||||||| :|.||.::|.|::||||||||||||||||||||||||:.:||:
Db   179  DMWANPHLFKTDVNGKATCIAGCPPDEFSVTGQLWGNPIYDWEAMDKDGYKKMWIERLRES 238

Qy   286  LKQCHLVRIDHFRGFEAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLG 345
          |||||||||||||||:||||||||||||||||||||||||||||::|.:||:|||||||
Db   239  FKIYDIVRIDHFRGFESYWEIPAGSDTAAPGEWVKGPGYKLFAAVKEELGELNIIAEDLG 298

Qy   346  VITPEVEALRDGFGFGPGMKILQFAFSGEDNAF-LPHNYPAHGNVVVYSGTHDNDTTLGWF 404
          |:|.|::.|| |  .|::|||||||||:.:  :  :|.:||||.:|:::|||||.||||
Db   299  FMTDEVIELRERTGFPGMKILQFAFNPEDESIDSPHLAPA--NSVMYTGTHDNNTVLGWY 356

Qy   405  RTAPEAERAFMRAYLARYGIRCLSEYE-VAGALIELAFKSPAKLAIVPLQDVLGLGPEAR 463
          |  ||||::|| ||.:|. :    ::|  ||:::|:||::  .|||.||||||::|:
Db   357  RN--EIDDA-TREYMARYTNR--KEYETVVHAMLRTVFSSVSFMAIATMQDLLELDEAAR  411

Qy   464  MNFPGRLGDNWAWRYAEGDLEPGLAAGLRALAEASQR 500
          ||||  |||:||||         .
Db   412  MNFPSTLGGNWSWRMTEDQLTPAVEEGLLDLTTIYRR 448
```

```
RESULT 4
US-10-424-599-170978
; Sequence 170978, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT:  La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION:  Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
;  SEQ ID NO 170978
;  LENGTH: 567
;  TYPE: PRT
;  ORGANISM: Glycine max
;  FEATURE:
;  NAME/KEY:  unsure
;  LOCATION: (1)..(567)
;  OTHER INFORMATION: unsure at all Xaa locations
;  FEATURE:
;  OTHER INFORMATION: Clone ID: PAT_MRT3847_125408C.1.pep
US-10-424-599-170978
Query Match         36.4%;  Score 989;  DB 4;  Length 567;
Best Local Similarity  44.2%;  Pred. No. 1.2e-83;
Matches  214;  Conservative  71;  Mismatches 181;  Indels  18;  Gaps  8;
```

```
Qy     2  QLQRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGP----TSYGDS  57
          :|:| .||||||:||.|.||||||||||||||||||||||||||||.  |   . :|||
Db    67  ELRRRCGILLHPTSFRGPYGIGDLGDEAFRFIDWLHHSGGCSVWQVLPLVPPGRKANEEGS 126

Qy    58  PYQSFSAFAGNPYLVDPEMLIEKGWLEQSEAPPPYTQRVDYGWLYQTRWPLLRRAFAGF 117
          ||||||||||| ||:|||:|||||||::|:||.|||||||||||||||||||||||||
Db   127  PYSGQDANCGNTLLISLEGLVEDGLLEKHSLPQPIDAERVNFSLVADVKDPLITKA-AER 185

Qy   118  RARASAQDKTRLEAFIEAERF--WLEDYALFMALKTRFDGKPWNEWSPELRDREPAALAR 175
          :|| ||:.|||||||||||||  ||||||||||||||||||||||:||||||||||||
Db   186  LISSEGELKTQLENFRRDPDISSWLEDAAYFAAIDDSLNTISWYNWPEPLRNRHLVALED 245
```

```
Db   176  AREELAEEVALYEWIQWLFYLLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFY 235

Qy   246  IYQQKRDFINVPIAQQFLFQRQWQKVHSYAQSKGISIMGDMPIYVGYHSADVWANKKQPL 305
          :|||||| ||||:|  |.|.||  ||.|||:||||.||||||||| |||||||||::|.
Db   236  LEADGNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRID 295

Qy   306  LNRKGPPLLVSGVPPDAFSETGQLWGSSPLYDWKAMEKDGYSWWVRRIRRAQNLFDEFRID 365
          |:||: |||.||:||||||||||||| ||||||  | ||| ||  : |||:|||:||||
Db   296  HFRGFEAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALR 355

Qy   366  HFRGFAGYWAVPSEAKVAMLGKW-XGPGISLFDTIFRAVGRINIIAEDLGVITEDVVQLR 424
          |||||.|||.||.:|::.|:.|| .||||||:|: ||||.|:|||||||||||:||||:|
Db   356  DGFGFPGMKILQPAFSGE-DNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAF 414

Qy   425  RSIGAPGMAVLQFGFGGDANNFHLPHNHEC--NQVVVTGTHDNDTIGGWWEALNQEEKSN 482
          |||||||| ||||.||||:||||:|:|:||  |||||:|||||||:|||||||||||||
Db   415  MRAYLARYGIRCLSE-YEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPPGRLGDN 473

Qy   483  VLSYLS------LNEGDEISWALIQRVLASVAQTAIIPMQDVLGLGNSARMNIPATQFGN 536
          |:|.|:      |:  ||.|| ||.|| |||||||||||||||||||:|||||.||:|.
Db   474  WAWR 477

Qy   537  WGWR 540
```

```
RESULT 5
US-10-425-114-57203
; Sequence 57203, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
;  APPLICANT:  Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION:  Nucleic Acid Molecules and Other Molecules Associated
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
;  SEQ ID NO 57203
;  LENGTH: 486
;  TYPE: PRT
;  ORGANISM: Zea mays
;  FEATURE:
;  OTHER INFORMATION: Clone ID: UC-ZMFLB73152G04_FLI.pep
US-10-425-114-57203
Query Match         33.0%;  Score 894.5;  DB 4;  Length 486;
Best Local Similarity  42.0%;  Pred. No. 7.3e-75;
Matches  196;  Conservative  63;  Mismatches 193;  Indels 15;  Gaps  7;
```

```
Qy    18  GRWGIGALGREAERFLDWLADAGARWWQVLPL---GPTSYGD-SPYQSFSAFAGNPYLVD  73
          |:||||.|:|||.||||||||||||||||||:   ||  :|| |||||||||||| |:|
Db     1  GPHGIGDLGDEALAFLHWLRDAGCTLWQVLPVPGRTAGEDGSPYSGQDANCGNTLLLS  60

Qy    74  PEMLIEKGWLEQSEAPPPYTQRVDYGWLYQTRWPLLRRAFAGFRARASAQDKTRLEAFI 133
          |:||||||||::|:||.|||||||||||||||||||||||.|||||||||.|||||||
Db    61  LEBLVKDGLLMSDELPDPVDTEYVEFDTVANLKEPLIAKA-AERLLLSHGELRTQYDCFK 119

Qy   134  EAERF--WLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAREELAEEVALYEWIQ 191
          |||||  |||||||||||||||||||||:|||||||||:||||||||||||||||||||
Db   120  KNPNISGWLEDAALFAAIDKSINAFSWYENWPEPLKNRHLGALEDIYEKQKDFIETFMAQQ 179

Qy   192  WLFYLLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADGNPTVVAGVPRD 251
          ||||||||||||||||||||||||||||| |||||||||:|||||||:||||||||||||
Db   180  FLFQRWQRIRKYAQKLGISIMGDMPIYVGYHSADVWANRKSFLLDKNGPPTFVSGVPPD 239

Qy   252  YFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGFEAYWEVPFGRP 311
          ||||||||||||||||||||||||||||||||||||||||| ::|||| :|||||||||
```

```
Db   240  AFSETGQLWNSPLYDWKAMEADCFSWWIKRIKRALDLYDEFRIDHFRGLAGFWAVPSDAK 299
          : :.: :|||||| .::  :.          ||||||||| :||||
Qy   312  NAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGFPGMKILQFAF-  370
           .:.          :. : ||||||||||||||     |||||||||||:
Db   300  VALVGSWRAGPRNSFFDTLFKAVGRIDIIAEDLGVITEDVVQLRKSIGAPGMAVLQFAFG 359
          :|||                      |||||||||| ::
Qy   371  SGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAYLARYGIRCLSEY 430
          .||:|  |||||| || ::|||||||||| |||     .
Db   360  GGSDNPHLPHNHEM--DQVVVTGTHDNDTVLGWWQNLPEEEKKIVIKYLPE-----AENI 412
          ||||||||   :||| |
Qy   431  EVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWR           477
          :|| || |||   :|  ||  ||      |||||| |  ||||  |
Db   413  DITWTLITAALSSVARTSVVTMQDILGLDSSARMNTPATQKGNWRWR          459
```

RESULT 6
US-10-336-597-5
; Sequence 5, Application US/10336597
; Publication No. US20030150021A1
; GENERAL INFORMATION:
;   APPLICANT: KREBBERS, ENNO
;   BROGLIE, KAREN E.
;   PEARLSTEIN, RICHARD W.
;   TITLE OF INVENTION: PLANT 4-(-GLUCANOTRANSFERASE
;   NUMBER OF SEQUENCES: 16
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: E. I. DUPONT DE NEMOURS AND COMPANY
;   STREET: 1007 MARKET STREET
;   CITY: WILMINGTON
;   STATE: DELAWARE
;   COUNTRY: UNITED STATES OF AMERICA
;   ZIP: 19898
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3.50 INCH DISKETTE
;   COMPUTER: IBM PC COMPATIBLE
;   OPERATING SYSTEM: MICROSOFT WINDOWS 95
;   SOFTWARE: MICROSOFT WORD FOR WINDOWS 95 (VER. 7.0)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/336,597
;   FILING DATE: 03-Jan-2003
;   CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/09/403,332
;   FILING DATE: 19-Oct-1999
;   ATTORNEY/AGENT INFORMATION:
;   NAME: MAJARIAN, WILLIAM R.
;   REGISTRATION NUMBER: 41,173
;   REFERENCE/DOCKET NUMBER: BB-1101-A
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 302-992-4926
;   TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 298 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
;   SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-336-597-5

  Query Match   28.6%; Score 777.5; DB 4; Length 298;
  Best Local Similarity   53.2%; Pred. No. 3.6e-64;
  Matches 156; Conservative   39; Mismatches   89; Indels   9; Gaps   6;

```
Qy   210  IQIIGDMPIFVAPDSSDVWANPQYFYLEADGNPTVVAGVPRDYFSETGQLWGNPLYRWDV 269
          |:|:|||||:||.||||:|||| : | .|  :|| .|| ||:||| ||:|||||::|| :
Db     3  IEIVGDMPIYVAEDSSDMWANPHLFKTDVNGKATCIAGCPPDEFSVIGQLWGNPIYDWEA 62
Qy   270  MERDNFAWCIARIRQSLKQCHLVRIDHFRGFEAXWEVFGRPNAVEGRWVKAPGEKLFAA 329
          |::|  :|  | :||.|:  : :||||||||||  : || ||||||:|||||.||||||
Db    63  MDKDGYKWWIERLRESFKIYDIVRIDHFRGFESSYWEIPAGSDTAAPGEWVKGPGYKLFAA 122
```

```
Db   330  VRAQLSDAPIIAEDLGVITPEVEALRDGFGFPGMKILQFAFSGEDNAF-LPHNYPAHGNV 388
          : :.: :|||||| |||| :|:: ||| ||||||||||||    ||:| |||| ||||:
Qy   123  VKEELGELNIIAEDLGFMTDEVIELRERTGFPGMKILQFAFNPEDESIDSPHLAPA--NS 180
          :: ||||||| | ::| |    | ||||| :| |. . |
Db   389  VVYSGTHDNDTTLGWFRTAPEAERAFMRAYLARYGIRCLSEYE-VAGALIELAFKSPAKL 447
          |: |||||| ||||: |    ||          |  :
Qy   181  VMYTGTHDNNTVLGWYRN--EIDDA-TREYMARYTNR--KEYETVVHAMLRTVFSSVSFM 235
          |: :|:||:||||   |  ::|    |:: :| ||  :
Db   448  AIVPLQDVLGLGPEARMNFPGRLGDNWAWRYAEGDLEPGLAAGLRALAEASQR 500
          || |:|  | || ||||||||  |||:| : : :| |:. | ||:::
Qy   236  AIATMQDLLELDEAARMNFPSTLGGNWSWRMTEDQLTPAVEEGLLDLTTIYRR 288
```

RESULT 7
US-10-425-115-325328
; Sequence 325328, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;   APPLICANT: La Rosa, Thomas J.
;   APPLICANT: Kovalic, David K.
;   APPLICANT: Zhou, Yihua
;   APPLICANT: Cao, Yongwei
;   TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;   TITLE OF INVENTION: Plants
;   FILE REFERENCE: 38-21(53222)B
;   CURRENT APPLICATION NUMBER: US/10/425,115
;   CURRENT FILING DATE: 2003-04-28
;   NUMBER OF SEQ ID NOS: 369326
  SEQ ID NO 325328
;   LENGTH: 486
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(486)
;   OTHER INFORMATION: unsure at all Xaa locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_59770C.1.pep
US-10-425-115-325328

  Query Match   28.3%; Score 767.5; DB 4; Length 486;
  Best Local Similarity   37.3%; Pred. No. 6.2e-63;
  Matches 185; Conservative   58; Mismatches 180; Indels   73; Gaps   11;

```
Qy    18  GRWGIGALGREAERFLDWLADAGARWWQVLPL---GPTSYGD-SPYQSFSAFAGNPYLVD 73
          | :||| ||..|| || |||||| .:||||||   |.:.| | ||| | .  .. ||
Db     1  GPHGIGDLGDEALAFLHWLRDAGCTLWQVLPLVPPGRTAGEDGSPYSGQDANCGNTLLLS 60
Qy    74  PEMLIEKGWLEQSEAPPYPTQRVDYGWLIYQTRWPLLRRAFGRARASAQDKTRLEAFI 133
           | |.: | |  ::      |:|   | | : | :|  .   :
Db    61  LEELVKDGLLMSDELPDPVDTEYVEFDTVANLKEPLIAKA------------- 100
Qy   134  EAERFWL------------------EDYALFMALKTRFDGKPWNEW--SPELRDREPAALARA 176
          .|| |.|                  || |:     ::| .
Db   101  -AERLLLSXESSEHNMTASKKIQIYLLVGTLGRQFSSLXFRRWXDNPHLPH--------- 150
Qy   177  REELAEEVALYEWIQWLFYLEWGQ------------TKAYA-ESKGIQIIGDMPIFVAF 222
          .   :  : :   : | . . .                :.:  :.||:|:|||||::
Db   151  NHEMDQVVYTXEHMITIQFLAGGKIYQRKXITWAKVXPRKYAXKTLRISIMGDMPLYAGY 210
Qy   223  DSSDVWANPQYFYLEADGNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARI 282
          .| .||||  : .::  ||.| ||.||.| |||||||| .||| |..:|.| .|  |.|
Db   211  HSAYVWANRKSFLLDKNGFPTFVSGVPPDAFSETGQLWNSPLYDWKAMEADCFSWWIKRI 270
Qy   283  RQSLKQCHLVRIDHFRGFEAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAE 342
          ::.|   . |||||||| | :||:| | |||||||||| |||||||||| |.| |:|||
Db   271  KRALDLYDEFRIDHFRGLAGFWAVPSDAKVALVGSWRAGPRNSFFDTLFKAVGRIDIIAE 330
Qy   343  DLGVITPEVEALRDGFGFPGMKILQFAF-SGEDNAFLPHNYPAHGNVVVYSGTHDNDTTL 401
          |||||| :|: ||| ||||||| |||||  .|.|| |||| ||| |:::|:|||||| |
Db   331  DLGVITEDVVQLRKSIGAPGMAVLQFAFPGGGSDNPHLPHNHEM--DQVVYTGTHDNDTVL 388
Qy   402  GWFRTAPEAERAFMRAYLARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPE 461
```

```
Db    389 GWWQNLPEEKKIVIKYLPE----AENIDITWTLITAALSSVARTSVVTMQDILGLDSS 443
          ||:::  ||::||:  ::||   |       ||   :  | ||:  :  ::||:|||
Qy    462 ARMNFPGRLGDNWAWR 477
          ||||  |||
Db    444 ARMNTPATQKGNWRWR 459
```

RESULT 8
US-10-336-597-6
; Sequence 6, Application US/10336597
; Publication No. US20030150021A1
; GENERAL INFORMATION:
;   APPLICANT: KREBBERS, ENNO
;   APPLICANT: BROGLIE, KAREN E.
;   APPLICANT: PEARLSTEIN, RICHARD W.
;   TITLE OF INVENTION: PLANT 4-(-GLUCANOTRANSFERASE
;   NUMBER OF SEQUENCES: 16
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: E. I. DUPONT DE NEMOURS AND COMPANY
;   STREET: 1007 MARKET STREET
;   CITY: WILMINGTON
;   STATE: DELAWARE
;   COUNTRY: UNITED STATES OF AMERICA
;   ZIP: 19898
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3.50 INCH DISKETTE
;   COMPUTER: IBM PC COMPATIBLE
;   OPERATING SYSTEM: MICROSOFT WINDOWS 95
;   SOFTWARE: MICROSOFT WORD FOR WINDOWS 95 (VER. 7.0)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/336,597
;   FILING DATE: 03-Jan-2003
;   CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/09/403,332
;   FILING DATE: 19-Oct-1999
;   ATTORNEY/AGENT INFORMATION:
;   NAME: MAJARIAN, WILLIAM R.
;   REGISTRATION NUMBER: 41,173
;   REFERENCE/DOCKET NUMBER: BB-1101-A
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 302-992-4926
;   TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 291 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
;   SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-336-597-6

  Query Match   25.6%; Score 694; DB 4; Length 291;
  Best Local Similarity   50.2%; Pred. No. 2.4e-56;
  Matches 136; Conservative 40; Mismatches 87; Indels 8; Gaps 3;

Qy   208 KGIQIIGDMPIFVAFDSSDVWANPQYFYLEADGNPTVVAGVPRDYFSETGQLWGNPLYRW 267
          |||:|||||| | : ::|||||| | :|:||||| ||||||||||:
Db   1 KGISIMGDMPIYVGYHSADVWANKKQFLLNRKGFPLIVSGVPPDAFSETGQLWGSPLYDW 60

Qy   268 DVMERDNFAWCIARIRQSLKQCHLVRIDHFRGFEAYWEVPFGRPNAVEGRWVKAPGEKLF 327
          |||||||||||||||||||:|||||||||||: |||||||||||||||||| ||||||||
Db   61 KAMEKDGFSWWVRRIQRATDLFDEFRIDHFRGFAGFWAVPSEEKIAILGRWKVGPGKPLF 120

Qy   328 AAVRAQLSDAPIIAEDLGVITPEVEALRDGFGFPGMKILQFAF-SGEDNAFLPHNYPAHG 386
          : :|||||||||||||||||||||||||||| ||||||||||| ||||:||:|||||
Db   121 DAILQAVGKINIIAEDLGVITEDVVQLRKSIEAPGMAVLQFAFGSDAENPHLPHNH--EQ 178

Qy   387 NVVVYSGTHDNDTTLGWFRTAPEAERAFMRAYLARYGIRCLSEYEVAGALIELAFKSPAK 446
          ||| |||||||||||||||||||||||||||||||||||||||||||: |:||||||:
Db   179 NQVVYTGTHDNDTIRGWWDTLPQEEKSNVLKYLSN----IEEEEISRGLIEGASSVAR 233
US-10-336-597-6
```

```
Qy    447 LAIVPLQDVLGLGPEARMNFPGRLGDNWAWR 477
          :||:|||:|||||||||||||:: |||
Db    234 IAIIPMQDVLGLGSDSRMNIPATQFGNWSWR 264
```

RESULT 9
US-10-437-963-138565
; Sequence 138565, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;   APPLICANT: La Rosa, Thomas J.
;   APPLICANT: Kovalic, David K.
;   APPLICANT: Zhou, Yihua
;   APPLICANT: Cao, Yongwei
;   APPLICANT: Wu, Wei
;   APPLICANT: Boukharov, Andrey A.
;   APPLICANT: Barbazuk, Brad
;   APPLICANT: Li, Ping
;   TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;   TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;   FILE REFERENCE: 38-21(53221)B
;   CURRENT APPLICATION NUMBER: US/10/437,963
;   CURRENT FILING DATE: 2003-05-14
;   NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 138565
;   LENGTH: 243
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_39941C.1.pep
US-10-437-963-138565

  Query Match   17.6%; Score 477; DB 4; Length 243;
  Best Local Similarity   45.3%; Pred. No. 4.6e-36;
  Matches 101; Conservative 25; Mismatches 89; Indels 8; Gaps 3;

Qy   256 TGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGFEAYWEVPFGRPNAVE 315
          ||| |||||:|| :||| ||||||||||||:|||||||||||| |||||||||||||
Db   1 TGQLWNSPLYDWKAMEAGGFEWWIKRINRALDLYDEFRIDHFRGLAGFWAVPSESKVALV 60

Qy   316 GRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGFPGMKILQFAF-SGED 374
          ||| :||| |||| :|||||||||||||||:|| :|| | || || ||||||| |:
Db   61 GSWRAGPRNAFFDALFKAVGRINIIAEDLGVITEDVVDLRKSIEAPGMAVLQFAFGGGSD 120

Qy   375 NAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAYLARYGIRCLSEYEVAG 434
          | || : |: ||||||||||||||| | || | |: || :   ::
Db   121 NPHLPHH--EFDQVVYTGTHDNDTVIGWWQTLPEEKQTVFKYLPE-----ANRTEISW 173

Qy   435 ALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWR 477
          || : ||| | :||:|||:||||| : |||: |||
Db   174 ALITAALSSVARTSMVTMQDILGLDSSARMNTPATQKGNWRWR 216
```

RESULT 10
US-09-841-132-518
; Sequence 518, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
;   APPLICANT: Bhatia, Ajay
;   APPLICANT: Skelky, Yasir A.W.
;   APPLICANT: Probst, Peter
;   TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
;   TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
;   FILE REFERENCE: 210121.469C8
;   CURRENT APPLICATION NUMBER: US/09/841,132
;   CURRENT FILING DATE: 2001-04-23
;   NUMBER OF SEQ ID NOS: 599
;   SOFTWARE: FastSEQ for Windows Version 3.0/4.0
; SEQ ID NO 518
;   LENGTH: 526
;   TYPE: PRT
;   ORGANISM: Chlamydia pneumoniae

US-09-841-132-518

Query Match   16.6%;   Score 451;   DB 3;   Length 526;
Best Local Similarity   26.8%;   Pred. No. 3.6e-33;
Matches 131;   Conservative   87;   Mismatches 212;   Indels   58;   Gaps   15;

```
Qy   21 GIGALGREAERFLD------WLADAGARWWQVLPLGPTSYGDSPYQSFSAFAGNPYLVDP 74
           |||              ::::|||||| :::|||||||| :::|||
Db   41 GIG------EFLDLIPLISWCQKQGFSVIQLLPLNDTGEDTSPYNSISSVALNPLFLSL 93

Qy   75 EML-------IEKGWLEQSEAPPPYPTQRVDYGWLYQTRWPLLRRAFAGFRARASAQDK 126
        :::       :: :||||:|||:  ::  ::|||: ::|||||||  ::||
Db   94 SSLPNIDTIPEVAKKLQDMHEL---CSTPSVSYTQVKEKKWAFLREYYQKC-CKSSLEGN 149

Qy  127 TRLEAFIEAERFWLEDYALFMALKTRFDGKPWNESPELRDREPAALARAAREELAEEVAL 186
          :::|||||:::||| :|||||: :::|||  :::|||| :::|||  :||::
Db  150 SNFSEFLESERYWLYPYGTFRAIKHHMHGEPINNWPKSLTDQE--NFPDLTKKFHDEVLF 207

Qy  187 YEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADGNPTVVA 246
        :::|:::||  :::|||||| :::|||  :::|||||::::||| :::||
Db  208 FSYLQFLCYQQLCEVKAYADQHHVLLKGDLPLILISKDSCDVWYFRDYF-----SSRSV 261

Qy  247 GVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGFEAYWEV 306
        |:  ::|:::|||: :::|||:  ::|||| :::|||  :::|||:  ::|||
Db  262 GAPPDLYNSEGQNWHLPIYNFSQLAKDDYIWWKERLRYAQNFYSVYRLDHIIGFFRLWI- 320

Qy  307 PFGRPNAVEGRWVK------APGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFG 359
         :::|||  :::|||||  :::|||| :::||||:  ::|||
Db  321 ---WDSSGRGRFIPDNPKDYIKQGTEILSTMLGASSMLP-IGEDLGIIPQDVKTTLTHLG 376

Qy  360 FPGMKILQFAFSGE-DNAFLPHNYPAHGNVVYSGTHDNDTTLGWFRTAPEAERAFMRAY 418
         :::|||||: ||| :::|||| :::||||::::||| :::|||
Db  377 ICGTRIPRWERNWESDSAFIPLKDYNPLSVTTLS-THDSDTFAQWMLNSPKEAKQFAKFL 435

Qy  419 LARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPE-------ARMNFPGRL 470
        : ::|||: :::|||| :::|||:  ::||||
Db  436 HLPFQKTLTTETQI--DILKLSHESASIFHINLFNDYLALCPDLVSKNLQRERINTPGTI 493

Qy  471 G-DNWAWR 477
        | :::||:
Db  494 SKKNWSYR 501
```

RESULT 11
US-10-289-762-347
; Sequence 347, Application US/10289762
; Publication No. US20040006218A1
; GENERAL INFORMATION:
;   APPLICANT: Griffais, R.
;   TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
;   TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis and treatment of infection
;   TITLE OF INVENTION: and treatment of infection
;   FILE REFERENCE: 9710-003-999
;   CURRENT APPLICATION NUMBER: US/10/289,762
;   CURRENT FILING DATE: 2003-03-27
;   NUMBER OF SEQ ID NOS: 6849
;   SEQ ID NO 347
;   LENGTH: 526
;   TYPE: PRT
;   ORGANISM: Chlamydia pneumoniae
US-10-289-762-347

Query Match   16.6%;   Score 451;   DB 4;   Length 526;
Best Local Similarity   26.8%;   Pred. No. 3.6e-33;
Matches 131;   Conservative   87;   Mismatches 212;   Indels   58;   Gaps   15;

```
Qy   21 GIGALGREAERFLD------WLADAGARWWQVLPLGPTSYGDSPYQSFSAFAGNPYLVDP 74
           |||              ::::|||||| :::|||||||| :::|||
Db   41 GIG------EFLDLIPLISWCQKQGFSVIQLLPLNDTGEDTSPYNSISSVALNPLFLSL 93

Qy   75 EML-------IEKGWLEQSEAPPPYPTQRVDYGWLYQTRWPLLRRAFAGFRARASAQDK 126
        :::       :: :||||:|||:  ::  ::|||: ::|||||||  ::||
Db   94 SSLPNIDTIPEVAKKLQDMHEL---CSTPSVSYTQVKEKKWAFLREYYQKC-CKSSLEGN 149

Qy  127 TRLEAFIEAERFWLEDYALFMALKTRFDGKPWNESPELRDREPAALARAAREELAEEVAL 186
          :::|||||:::||| :|||||: :::|||  :::|||| :::|||  :||::
Db  150 SNFSEFLESERYWLYPYGTFRAIKHHMHGEPINNWPKSLTDQE--NFPDLTKKFHDEVLF 207

Qy  187 YEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADGNPTVVA 246
        :::|:::||  :::|||||| :::|||  :::|||||::::||| :::||
Db  208 FSYLQFLCYQQLCEVKAYADQHHVLLKGDLPLILISKDSCDVWYFRDYF-----SSRSV 261

Qy  247 GVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGFEAYWEV 306
        |:  ::|:::|||: :::|||:  ::|||| :::|||  :::|||:  ::|||
Db  262 GAPPDLYNSEGQNWHLPIYNFSQLAKDDYIWWKERLRYAQNFYSVYRLDHIIGFFRLWI- 320

Qy  307 PFGRPNAVEGRWVK------APGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFG 359
         :::|||  :::|||||  :::|||| :::||||:  ::|||
Db  321 ---WDSSGRGRFIPDNPKDYIKQGTEILSTMLGASSMLP-IGEDLGIIPQDVKTTLTHLG 376

Qy  360 FPGMKILQFAFSGE-DNAFLPHNYPAHGNVVYSGTHDNDTTLGWFRTAPEAERAFMRAY 418
         :::|||||: ||| :::|||| :::||||::::||| :::|||
Db  377 ICGTRIPRWERNWESDSAFIPLKDYNPLSVTTLS-THDSDTFAQWMLNSPKEAKQFAKFL 435

Qy  419 LARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPE-------ARMNFPGRL 470
        : ::|||: :::|||| :::|||:  ::||||
Db  436 HLPFQKTLTTETQI--DILKLSHESASIFHINLFNDYLALCPDLVSKNLQRERINTPGTI 493

Qy  471 G-DNWAWR 477
        | :::||:
Db  494 SKKNWSYR 501
```

RESULT 12
US-10-872-155-518
; Sequence 518, Application US/10872155
; Publication No. US20040234536A1
; GENERAL INFORMATION:
;   APPLICANT: Bhatia, Ajay
;   APPLICANT: Skeiky, Yasir A.W.
;   APPLICANT: Probst, Peter
;   TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND
;   TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
;   FILE REFERENCE: 210121.469C10
;   CURRENT APPLICATION NUMBER: US/10/872,155
;   CURRENT FILING DATE: 2004-06-18
;   PRIOR APPLICATION NUMBER: 09/620,412
;   PRIOR FILING DATE: 2000-07-20
;   PRIOR APPLICATION NUMBER: 09/598,419
;   PRIOR FILING DATE: 2000-06-20
;   PRIOR APPLICATION NUMBER: 09/556,877
;   PRIOR FILING DATE: 2000-04-19
;   PRIOR APPLICATION NUMBER: 09/454,684
;   PRIOR FILING DATE: 1999-12-03
;   PRIOR APPLICATION NUMBER: 09/426,571
;   PRIOR FILING DATE: 1999-10-01
;   PRIOR APPLICATION NUMBER: 09/410,568
;   PRIOR FILING DATE: 1999-10-22
;   PRIOR APPLICATION NUMBER: 09/410,568
;   PRIOR FILING DATE: 1999-10-01
;   PRIOR APPLICATION NUMBER: 09/288,594
;   PRIOR FILING DATE: 1999-04-08
;   PRIOR APPLICATION NUMBER: 09/208,277
;   PRIOR FILING DATE: 1998-12-08
;   NUMBER OF SEQ ID NOS: 599
;   SOFTWARE: FastSEQ for Windows Version 3.0/4.0
SEQ ID NO 518
;   LENGTH: 526
;   TYPE: PRT
;   ORGANISM: Chlamydia pneumoniae
US-10-872-155-518

Query Match   16.6%;   Score 451;   DB 5;   Length 526;
Best Local Similarity   26.8%;   Pred. No. 3.6e-33;
Matches 131;   Conservative   87;   Mismatches 212;   Indels   58;   Gaps   15;

```
Qy   21 GIGALGREAERFLD------WLADAGARWWQVLPLGPTSYGDSPYQSFSAFAGNPYLVDP 74
           |||              ::::|||||| :::||||
```

```
Db   234  SCLIKGDIPILISKDSCDVWFYRHYF------          :  :.:  SSSESVGAPPDLYNAEGQNWHLPICNMK  287
                ||  |:  ||||  |||                                  :  ||  .||
Qy   269  VMERDNFAWCIARIRQSLKQCHLVRIDHFRGFEAYW---EVPFGRPNAVEGRWVKAPGEK  325
                :.:|  |:  |:.                     ||  |:.:
Db   288  TLQQDNYLWWKERLRYAENFYSLYRLDHVVGLPRFWVWDESGCGRFEPHDPKNYLAQGQD  347
                ||  ||  ::                ::  |||
Qy   326  LFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGPGMKILQFAFSGEDN-AFLPHNYPA   384
                :  ::|  ||||                  |:.:::  ||
Db   348  ILSHLLTSSSMLP-IGEDLGTIPSDVXRMLESFAVCGTRIPRWERNWEGNGAYTP--FDQ  404
                ||  |||                              |:  |:
Qy   385  HGNVVVYS-GTHDNDTTLGWFRTAPEAERAFMR--          AYLARYGIRCLSEYEVAGALIEL  439
                :  |:  :||:  |:  |:.                         ||  |||:
Db   405  YDPLSVTSLSTHDSSTLASWNKESPQESKLFAQFLGLPYSSTLSLHNHTE----ILKL   458
                ||  ||                      ::  |||
Qy   440  AFKSPAKLAIVPLQDVLGLGPE------          ARMNFPGRLG-DNWAWRYAEGDLEPGLAAG  490
                :  |:  |:  ||||                        ||||||||
Db   459  SHKTSSIFRINLINDYIALFPDLISKTPRYERINLPGTISKNNWVYRVKPSIEDLSSHSK  518
                ::  ||  :  :.                :|:||:||||
Qy   491  LRALAEA  497
                |:||||
Db   519  LNSLLEA  525
```

RESULT 14
US-10-872-155-588
; Sequence 588, Application US/10872155
; Publication No. US20040234536A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C10
; CURRENT APPLICATION NUMBER: US/10/872,155
; CURRENT FILING DATE: 2004-06-18
; PRIOR APPLICATION NUMBER: 09/620,412
; PRIOR FILING DATE: 2000-07-20
; PRIOR APPLICATION NUMBER: 09/598,419
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/556,877
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: 09/454,684
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: 09/426,571
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: 09/410,568
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: 09/288,594
; PRIOR FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: 09/208,277
; PRIOR FILING DATE: 1998-12-08
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSEQ for Windows Version 3.0/4.0
; SEQ ID NO 588
; LENGTH: 527
; TYPE: PRT
; ORGANISM: C. Trachomatis D serovar
US-10-872-155-588

Query Match   15.9%; Score 430.5; DB 5; Length 527;
Best Local Similarity   25.9%; Pred. No. 3.1e-31;
Matches 126; Conservative   89; Mismatches 233; Indels   39; Gaps   12;

```
Qy    33  LDWLADAGARWWQVLPLGPTSYGDSPYQSFSAFAGNPYLVDPEMLIEKGWLEQSEAPPPY   92
                ||  :|      :  ||  ||||  ||  :  |||||              |:
Db    56  IDWCISCGFQILQILPINDTGSCSSPYNSISSIALNPLHLSISALPYKEEVPAAETRIRE  115
                :  ||                              ::  :  |||
Qy    93  PTQRVDYGWLYQTRWPLLRRAFAGFRARASAQ----DKTRLEAFIEAERFWLEDYALFMA  148
                :|      |  ||                    ||||        :  |||||
Db   116  MQQLSQLPQVHYEKVRSMKRDFFQEYYRVCKQKKLTDHPDFYAFCEQEKYWLHPYALFRS  175
```

```
Db    41  GIG------EFLDLIPLISWCQKQGFSVIQLLPLNDTGEDTSPYNSISSVALNPLFLSL   93
                |||                              |  ||  ||||  ||||  :  |||
Qy    75  EML--------IEKGWLEQSEAPPPYPTQRVDYGWLYQTRWPLLRRAFAGFRARASAQDK  126
                |:                                |  |:
Db    94  SSLPNIDTIPEVAKKLQDMEHL---CSTPSVSYTQVKEKKWAFLREYYQKC-CKSSLEGN  149
                :  |:  :  |  ||                          ::  :  ||
Qy   127  TRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAREELABEVAL  186
                :  ||  :  :|||||              :  |||
Db   150  SNFSEFLESSERYWLYPYGTFRAIKHHMHGEPINNWPKSLTDQE--NFPDLTKKFHDEVLF  207
                :  ||  :|||                          :|  |||
Qy   187  YEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADGNPTVVA  246
                :  |||||                    :|  |:  |||
Db   208  FSYLQFICYQQLCEVKAYADQHHVLLKGDLPILISKDSCDVWYFRDYF------SSSRSV  261
                |:  ||  ||                    ::  ||||
Qy   247  GVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGFEAYWEV  306
                :||  ||  |:|                |:  ||  |:.
Db   262  GAPPDLYNSEGQNWHLPIYNFSQLAKDDYIWWKERLRYAQNFYSVRLDHIIGFFRLWI-  320
                :|  |:  |:|                ::  ||  ::
Qy   307  PFGRPNAVEGRWVK------APGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFG   359
                |  ||  :  |:                      ::  ||  ||||
Db   321  ---WDSSGRGRFIPDNPKDYIKQGTEILSTMLGASSMLP-IGEDLGIIPQDVKTTLTHLG  376
                :  ::  :|                    :  ||||
Qy   360  FPGMKILQFAFSGE-DNAFLPHNYPAHGNVVYYSGTHDNDTTLGWFRTAPEAERAFMRAY  418
                ||||:  |:  :  :::                :||:  |:  |:.  |||
Db   377  ICGTRIPRWERNWESDSAFIPLKDYNPLSVTTLS-THDSDTFAQWMLNSPKEAKQFAKFL  435
                :  ||  :  :.                :|:  |:  :  |||
Qy   419  LARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPE------ARMNFPGRL   470
                |  ||:  |:                            :|  |:  ||||||
Db   436  HLPFQKTLTTETQI--DILKLSHESASIFHINLFNDYLALCPDLVSKNLQRERINTPGTI  493
                |:  |:                ::  :  ||
Qy   471  G-DNWAWR  477
                |:||  |:
Db   494  SKKNWSYR  501
```

RESULT 13
US-09-841-132-588
; Sequence 588, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSEQ for Windows Version 3.0/4.0
; SEQ ID NO 588
; LENGTH: 527
; TYPE: PRT
; ORGANISM: C. Trachomatis D serovar
US-09-841-132-588

Query Match   15.9%; Score 430.5; DB 3; Length 527;
Best Local Similarity   25.9%; Pred. No. 3.1e-31;
Matches 126; Conservative   89; Mismatches 233; Indels   39; Gaps   12;

```
Qy    33  LDWLADAGARWWQVLPLGPTSYGDSPYQSFSAFAGNPYLVDPEMLIEKGWLEQSEAPPPY   92
                ||  :|      :  ||  ||||  ||  :  |||||              |:
Db    56  IDWCISCGFQILQILPINDTGSCSSPYNSISSIALNPLHLSISALPYKEEVPAAETRIRE  115
                :  ||                              ::  :  |||
Qy    93  PTQRVDYGWLYQTRWPLLRRAFAGFRARASAQ----DKTRLEAFIEAERFWLEDYALFMA  148
                :|      |  ||                    ||||        :  |||||
Db   116  MQQLSQLPQVHYEKVRSMKRDFFQEYYRVCKQKKLTDHPDFYAFCEQEKYWLHPYALFRS  175
                :  ||                              ::  :  :  |||
Qy   149  LKTRFDGKPWNEWSPELRDREPAALARAREELAEEVALYEWIQWLFYLEWGQTKAYAESK  208
                :  ||  :  :.                :|:  ||
Db   176  IREHLDNLPINHWPTTYTDL--SQTEHERTFAEDIQFHSYLQYLCFQQMTQVREHANCK   233
                :  ||                              ::  :||
Qy   209  GIQIIGDMPIFVAFDSSDVWANPQYFYLEADGNPTVVAGVPRDYFSETGQLWGNPLYRWD  268
```

```
Qy    149  LKTRFDGKPWNEWSPELRDREPAALARABELAEEVALYEWIQWLFYLEWGQTKAYAESK  208
                ::  ||     |  |     ::   |||:  :|: :|  :: :|:| |  |
Db    176  IREHLDNLPINHWPTTYTDL--SQITEHERTFAEDIQFHSYLQYLCFQQMTQVREHANCK  233

Qy    209  GIQIIGDMPIFVAFDSSDVWANPQYFYLEADGNPTVVAGVPRDYFSETGQLWGNPLYRWD  268
                ||:|:|| :: ||||        |       ||     |  |  ||:|  |  |
Db    234  SCLIKGDIPILISKDSCDVWFYRHYF------SSSESVGAPPDLYNAEGQNWHLPICNMK  287

Qy    269  VMERDNFAWCIARIRQSLKQCHLVRIDHFRGFEAYW---EVPFGRPNAVEGRWVKAPGEK  325
                ::||:|:   ::   :         ||||       ||      ||    |:
Db    288  TLQQDNVLWWKERLRYAENFYSLYRLDHVGLFRFWVWDESGCGRFEPHDPKNYLAQGQD  347

Qy    326  LFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGFPGMKILQFAFSGEDN-AFLPHNYPA  384
             : :       |  ||||||||||    :  ::      ||     ::|  |||||||
Db    348  ILSHLLTSSSMLP-IGEDLGTIPSDVKRMLESFAVCGTRIPRWERNWEGNGAYTP--FDQ  404

Qy    385  HGNVVYS-GTHDNDTTLGWFRTAPEAERAFMR----AYLARYGIRCLSEYEVAGALIEL  439
            :  :||   :  ||:|::  ::|: :::  |          |  |  |:   :|:|:|
Db    405  YDPLSVTSLSTHDSSTLASWWKESPQESKLFAQFLGLPYSSTLSLHNHTE------ILKL  458

Qy    440  AFKSPAKLAIVPLQDVLGLGPE---------ARMNFPGRLG-DNWAWRYAEGDLEPGLAAG  490
             |  : :   |:   |  :  |            :|    :|    |:| ||  :|  ::
Db    459  SHKTSSIFRINLINDYLALFPDLISKTPRYERINLPGTISKNNWVYRVKPSIEDLSSHSK  518

Qy    491  LRALAEA  497
            | | |  ||
Db    519  LNSLLEA  525


RESULT 15
US-10-767-701-51019
; Sequence 51019, Application US/10767701
; Publication No. US2004017268A1
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT: Zhou, Yihua
;  APPLICANT: Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE: 2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 51019
;  LENGTH: 179
;  TYPE: PRT
;  ORGANISM: Sorghum bicolor
;  FEATURE:
;  OTHER INFORMATION: Clone ID: LIB3480-052-P1-K1-G9.pep
US-10-767-701-51019

Query Match          15.2%;  Score 412.5;  DB 4;  Length 179;
Best Local Similarity 49.4%;  Pred. No. 3.5e-30;
Matches    83;  Conservative  18;  Mismatches  64;  Indels  3;  Gaps  2;

Qy    253  FSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGFEAYWEVPFGRPN  312
            ||||||||||:|||||  |||||:|:|: :| ||||:   |||:|||||   | | |||
Db      2  FSETGQLWNSPLYDWKAMEADGFSWWTKRIKRALDLYDEFRIDHFRGLAGFWAVPSDAKV  61

Qy    313  AVEGRWVKAPGEKLPAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGFPGMKILQFAF-S  371
            ||  |  ||  | |  ||| ||  ||||||||||||  :|||:||||||||||:  ||||
Db     62  ALVGSWRAGPRNSFFDALFKAVGRINIIAEDLGVITEDVVQLRKSIGAPGMAVLQFAFGG  121

Qy    372  GEDNAFLPHNYPAHGNVVYSGTHDNDTTLGWFRTAPEAERAFMRAYL  419
            |||||:|:|||||  :  |||||||||||||||:   :|| ||:  ||
Db    122  GSDNPHLPHNHEM--DQVVYTGTHDNDTVLGWWQNLPEEKKIVIKYL  167


Search completed: January 10, 2006, 14:24:33
Job time : 116 secs
```

OM protein - protein search, using sw model

```
            Copyright  GenCore version 5.1.6
            Copyright (c) 1993 - 2006 Compugen Ltd.

Run on:     January 10, 2006, 14:13:34 ; Search time 29 Seconds
                        (without alignments)
                        146.735 Million cell updates/sec

Title:         US-10-618-976-2
Perfect score: 2714
Sequence:      1 MQLQRAFGILHPTSFPGRW.........DLEPGLAAGLRALARASQRA 501

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched:   61141 seqs, 8493638 residues

Total number of hits satisfying chosen parameters:    61141

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    Published Applications AA_New:*
        1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep;*
        2: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep;*
        3: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep;*
        4: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep;*
        5: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep;*
        6: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep;*
        7: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep;*
        8: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep;*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	100	3.7	658	6	US-10-873-528-17	Sequence 17, Appl
2	100	3.7	677	6	US-10-873-528-155	Sequence 155, App
3	90	3.3	2644	6	US-10-770-726-45	Sequence 45, Appl
4	89	3.3	633	6	US-10-467-657-730	Sequence 730, App
5	89	3.3	1766	7	US-11-075-185-10	Sequence 10, Appl
6	88	3.2	620	7	US-11-055-822-460	Sequence 460, App
7	88	3.2	620	7	US-11-055-822-702	Sequence 702, App
8	87	3.2	538	7	US-11-167-856-20	Sequence 20, Appl
9	86	3.2	512	7	US-11-012-762-64	Sequence 64, Appl
10	85.5	3.2	695	7	US-11-038-284-34	Sequence 34, Appl
11	85	3.1	6893	7	US-11-205-109-14	Sequence 14, Appl
12	84.5	3.1	355	6	US-10-454-437-102	Sequence 102, App
13	84.5	3.1	555	6	US-10-454-437-100	Sequence 100, App
14	84	3.1	379	6	US-10-858-730-22	Sequence 22, Appl
15	82	3.0	715	6	US-10-467-657-70	Sequence 70, Appl
16	82	3.0	715	6	US-10-467-657-3672	Sequence 3672, Ap
17	81	3.0	294	6	US-10-467-657-934	Sequence 934, App
18	80	2.9	395	7	US-11-075-185-13	Sequence 13, Appl
19	80	2.9	402	6	US-10-454-437-282	Sequence 282, App
20	80	2.9	486	6	US-10-131-826A-58	Sequence 58, Appl
21	80	2.9	615	6	US-10-821-234-873	Sequence 873, App
22	80	2.9	801	6	US-10-454-437-278	Sequence 278, App
23	79.5	2.9	554	6	US-10-850-816-6	Sequence 6, Appli
24	79.5	2.9	582	7	US-11-090-439-58	Sequence 58, Appl
25	79.5	2.9	625	7	US-11-055-822-522	Sequence 522, App
26	79	2.9	561	6	US-10-454-437-330	Sequence 330, App
27	79	2.9	7968	7	US-11-143-980-49	Sequence 49, Appl
28	78.5	2.9	379	6	US-10-858-730-285	Sequence 285, App
29	78.5	2.9	589	6	US-10-821-234-1687	Sequence 1687, Ap
30	78	2.9	496	7	US-11-186-284-103	Sequence 103, App
31	77.5	2.9	379	6	US-10-858-730-24	Sequence 24, Appl
32	77.5	2.9	379	6	US-10-858-730-281	Sequence 281, App
33	77.5	2.9	379	6	US-10-858-730-290	Sequence 290, App
34	77.5	2.9	600	7	US-11-205-109-39	Sequence 39, Appl
35	77.5	2.9	777	6	US-10-821-234-1658	Sequence 1658, Ap
36	77.5	2.9	4655	6	US-10-995-561-556	Sequence 556, App
37	76	2.8	453	6	US-10-878-556A-141	Sequence 141, App
38	76	2.8	568	7	US-11-055-822-240	Sequence 240, App
39	76	2.8	568	7	US-11-055-822-1134	Sequence 1134, Ap
40	75.5	2.8	239	7	US-11-165-211-38	Sequence 38, Appl
41	75.5	2.8	575	6	US-11-165-226-48	Sequence 48, Appl
42	75.5	2.8	626	6	US-10-467-657-1456	Sequence 1456, Ap
43	75.5	2.8	626	6	US-10-512-184-49	Sequence 49, Appl
44	75.5	2.8	7102	7	US-11-143-980-48	Sequence 48, Appl
45	75	2.8	1892	7	US-11-075-185-6	Sequence 6, Appli

ALIGNMENTS

```
RESULT 1
US-10-873-528-17
; Sequence 17, Application US/10873528
; Publication No. US2005027681A1
; GENERAL INFORMATION:
;  APPLICANT: Microbial Technics Limited
;  APPLICANT: Gilbert, Christophe FG
;  APPLICANT: Hansbro, Philip M
;  TITLE OF INVENTION: Proteins
;  FILE REFERENCE: PWC/P21129WO
;  CURRENT APPLICATION NUMBER: US/10/873,528
;  CURRENT FILING DATE: 2004-06-23
;  PRIOR APPLICATION NUMBER: US/09/769,787
;  PRIOR APPLICATION NUMBER: 2001-01-26
;  PRIOR APPLICATION NUMBER: GB 9816337.1
;  PRIOR FILING DATE: 1998-03-27
;  PRIOR APPLICATION NUMBER: US 60/125164
;  PRIOR FILING DATE: 1999-03-19
;  NUMBER OF SEQ ID NOS: 388
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
;  LENGTH: 658
;  TYPE: PRT
;  ORGANISM: Streptococcus pneumoniae
US-10-873-528-17

Query Match             3.7%;  Score 100; DB 6; Length 658;
Best Local Similarity  18.9%;  Pred. No. 0.19;
Matches  103; Conservative  63; Mismatches 155; Indels 224; Gaps 28;

Qy    22 IGALGREAERFLDWLAD----------------AGARWWQVLPLGPTSYGDSPYQSFSAFA 66
             :|| :|  :  |:              :||.| ||  |  :  |   |
Db   116 VGATG--AKVIEDWVYDSQYDAWFYIKADGQHAEKEWLQI------KGKDY------FK 161

Qy    67 GNPYLVDPEMLIEKGWLEQSEAPPYPYPTQRVDYGWLYQTRWPLLRRAFAGFRARASAQDK 126
                                                 ||:| ||| :|  :||  |||
Db   162 SGGYLLTSQW--INQAYVNASGA----------KVQQGWLFDKQY--QSWFYIKENGNYADK 209

Qy   127 TRLEAFIE----------------------AERFWLEDYALFMALKTRFDGK-------PW- 158
            |:|  :|                    |:  |::   :||  ||
Db   210 ---EWIFENGHYYYLKSGGYMAANEWIWDKESWFYLK--FDGKMAEKEWVYDSHSQAWYY 264

Qy   159 --------NEWSPELRDREPAALARAREELAEEVALYEWIQWLFYLEWGQTKAYAESKG 209
                                    :  ::  ||  :|  ::|
Db   265 FKSGGYMTANEW---IWDKESWFYLKSDGKIAEK----EWV----YDSHSQAWYFKSGG 313

Qy   210 IQIIGDMPIFVAFDSSDVWANPQYFYLEADG-------------------NP 242
            :              :|||:|:||
```

```
Db  314 YMTANEW---------IWDKESWFYLKSDGKIAEKEWVYDSHSQAWYFKSGGYMAKNE 363
Qy  243 TVVA--------------------------------GVPRDYFSETGQLWGNPLYRW 267
           ||
Db  364 TVDGYQLGSDGKWLGGKTTNENAAYYQVVPTANVYDSDGEKLSYISQGSVVW------ 416
Qy  268 DVMERD------NFAWCIARIRQSLKQCHLVRIDHFRGFEAYWEVPFGRPNAVEGRWVKA 321
           :::
Db  417 --LDKDRKSDDKRLAITISGLSGYMKTEDLQALDASKDFIPYYE-------- 459
Qy  322 PGEKLFAAVRAQLSDAPIIA--EDLGVITPEVEALRDGFGFPGMKILQFAFSGEDNAFL- 378
           :::
Db  460 DGHRFYHYV-AQNASIPVASHLSDMEVGKKYYSA--DGLHFDGFKL-------ENPFLF 508
Qy  379 -----PHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAYLARYGIRCLSEYEVA 433
           :|:|:  :|: |:  :|:| :::
Db  509 KDLTEATNYSAEELDKVFSLLNINNSLL----------ENKGATFKEAEEHYHINAL--YLLA 559
Qy  434 GALIE 438
           :|
Db  560 HSALE 564
```

RESULT 2
US-10-873-528-155
; Sequence 155, Application US/10873528
; Publication No. US20050276814A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/10/873,528
; CURRENT FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 155
; LENGTH: 677
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-155

Query Match   3.7%;   Score 100; DB 6; Length 677;
Best Local Similarity   18.9%;   Pred. No. 0.2;
Matches 103; Conservative 63; Mismatches 155; Indels 224; Gaps 28;

```
Qy   22 IGALGREAERFLDWLAD----------AGARWWQVLPLGPTSYGDSPYQSFSAFA 66
           ||
Db  135 VGATG--AKVIEDWVDSQYDAWFYIKADGQHAEKEWLQI-------KGKDYY------FK 180
Qy   67 GNPYLVDPEMLIEKGWLEQSEAPPPYPTQRVDYGWLYQTRWPLLRRAFAGFRARASAQDK 126
           |||:  ::  :||
Db  181 SGGYLLTSQW-INQAYVNASGA-------KVQQGWLFDKQY----QSWFYIKENGNYADK 228
Qy  127 TRLEAFIE--------AERFWLEDYALFMALKTRFDGK-----------PW--- 158
           ||
Db  229 ---EWIFENGHYYLKSGGYMAANEWIWDKESWFYLK--FDGKMAEKEWYDSHSQAWYY 283
Qy  159 --------NEWSPELRDREPAALARAREELAEEVALYEWIQWLFYLEWGQTKAYAESKG 209
           :||
Db  284 FKSGGYMTANEW---IWDKESWFYLKSDGKIAEK----EWV----YDSHSQAWYYFKSGG 332
Qy  210 IQIIGDMPIFVAFDSSDVWANPQYFYLEADG----------NP 242
           :||||
Db  333 YMTANEW---------IWDKESWFYLKSDGKIAEKEWVYDSHSQAWYYFKSGGYMAKNE 382
```

```
Qy  243 TVVA--------------------------------GVPRDYFSETGQLWGNPLYRW 267
           ||
Db  383 TVDGYQLGSDGKWLGGKTTNENAAYYQVVPTANVYDSDGEKLSYISQGSVVW----- 435
Qy  268 DVMERD------NFAWCIARIRQSLKQCHLVRIDHFRGFEAYWEVPFGRPNAVEGRWVKA 321
           :::
Db  436 --LDKDRKSDDKRLAITISGLSGYMKTEDLQALDASKDFIPYYE--------S 478
Qy  322 PGEKLFAAVRAQLSDAPIIA--EDLGVITPEVEALRDGFGFPGMKILQFAFSGEDNAFL- 378
           :::
Db  479 DGHRFYHYV-AQNASIPVASHLSDMEVGKKYYSA--DGLHFDGFKL-------ENPFLF 527
Qy  379 -----PHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAYLARYGIRCLSEYEVA 433
           :|:|:  :|: |:  :|:| :::
Db  528 KDLTEATNYSAEELDKVFSLLNINNSLL----------ENKGATFKEAEEHYHINAL--YLLA 578
Qy  434 GALIE 438
           :|
Db  579 HSALE 583
```

RESULT 3
US-10-770-726-45
; Sequence 45, Application US/10770726
; Publication No. US20050266409A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Brown, Eugene
; APPLICANT: Liu, Wei
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING, PREVENTING, AND TREATIN
; TITLE OF INVENTION: CANCERS
; FILE REFERENCE: AM101079 (031896-010000)
; CURRENT APPLICATION NUMBER: US/10/770,726
; CURRENT FILING DATE: 2004-02-04
; NUMBER OF SEQ ID NOS: 48640
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 45
; LENGTH: 2644
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-770-726-45

Query Match   3.3%;   Score 90; DB 6; Length 2644;
Best Local Similarity   20.7%;   Pred. No. 10;
Matches   73; Conservative 39; Mismatches 39; Indels 102; Gaps 16;

```
Qy   18 GRWGIGALGREAERPLDWLADAGARWWQVLPLGPTSYGDSPYQSFSAFAGNPYLVDPEML 77
           ||  |  ::  :    :  |    | |:  :  |    |
Db 1805 GQLLLSAKKRDITAFYDSLKIVRAB--QIVPLSAASFERGSYQ----RGYEYIVRLHML 1857
Qy   78 IEKGWLEQSEAP------PPYPTQRVDYGWL--------YQTRWPL--LRRAF 114
           ||        |    :|  |  |: ||         |  | |  ||||
Db 1858 CE---LEHSIKPLFQHSPGDSSQEDSLNWVARLEMTQNSYRAKEPILALRRALLSLNKRP 1914
Qy  115 ----------------------------AGFRARASAQDKTRLEAFIEAERFWLEDYALFMALK 150
                                       :|  | |    | ||:  | |  ||
Db 1915 DYNEMVGECWLQSARVARKAGHHQTAYNALLNAGESRLAELYVERAKWLWSKGDVHQALI 1974
Qy  151 TRFDGK----PWNEWSPE----LRDREPAALARAREELAE------------EVALYE 188
           ||          :|       |:|  | ||  |             ||| |
Db 1975 VLQKGVELCFPENETPPEGKNMLIHGRAMLLVGRFMEETANFESNAIMKKYKDVTACLPE 2034
Qy  189 WIQWLFYLE--WGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADGNPTFVA 246
           |:| | ||   || |:|    ||:|:|||||||||||  | :   ||||:|| |
Db 2035 WEDGHFYLAKYYDKLMPMVTDNKMEKQGDLIRYIVLHFG---RSLQY---GNQFIYQ 2085
Qy  247 GVPRDYFSETGQLW--GNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDH 296
           |||||||||||||   :|||  |    |  |    |    | |||||:|
Db 2086 SMPRML----------TLWLDYGTKAYEWEKAGRSD---RVQMRNDLGKINKVITH 2129
```

RESULT 4
US-10-467-657-730

```
; Sequence 730, Application US/10467657
; Publication No. US2005026O581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: Seqwin99, version 1.04
; SEQ ID NO 730
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-730

Query Match            3.3%;  Score 89;  DB 6;  Length 633;
Best Local Similarity 18.7%;  Pred. No. 1.8;
Matches 100; Conservative 64; Mismatches 180; Indels 192; Gaps 30;

Qy   47 LPLGPTS--YGDSPYQSFSAFAGNPYLVDPEMLIEKGWLEQSEAPPYPTQRVDYGWLYQ 104
Db   70 IPVYDTSGAYGD-PAAHIDLKQGLPH-------IRTAWLDERGDTEILPKLSSEYG-IER 120

Qy  105 TRWPLL----------RRAFAG-----FRAR----------ASAQDKTRLEAFIEAER 137
Db  121 AHDPKTAHLRFNQITRPRRAKAGRNVTQLHYARQGIITPEMEFAAIRERMKLD-----EL 175

Qy  138 FWLEDYALFMA----------------LKIRFD----------GK--PWNEWSPELRDREPA 171
Db  176 FRRPEYAKLLKQHTGQSFGANIPTRPDQITPEFVRQEIAAGRAIIPANINHPEL---EPM 232

Qy  172 ALAR-------------AREELAEEVALYEWIQWLFYLEWGQTKAYAESKGIQI------ 212
Db  233 IIGRNFRVKINGNLGNSAVTSSLTEEVEKMVW-----SLRWGADTIMDLSTGAHIHETRE 287

Qy  213 -----------IGDMPIFVAFDSSD-------VW-----------ANPQYFYLEADGNPTVV 245
Db  288 WIIRNAPVPIGTVPIYQTLEKTGGIAEDLTWDLFRDTLIEQAEQGVDYF-------TIH 339

Qy  246 AGVPRDYFSET-GQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRI-DHFRGFEAY 303
Db  340 AGVLLRYVPMTANRLTGIVSRGGSIMAK----WCLAHHRENFLYTHFDEICEIMKAYDVS 395

Qy  304 WEVPFG-RPNAV-------------EGRW--------VKAPGEKLFAAVRAQLSD 336
Db  396 FSLGDGLRPGCIADANDESQPAELHTLGELTDKAWKHDVQVMIEGPGHVPLQRVKENMTE 455

Qy  337 ------------APIIAEDLGVITPEVEALRDG-----FGFPGMKILQFAFSGEDNAFL 378
Db  456 ELQHCFEAPFYTLGPLVTD---IAPGYDHITSGIGAANIGWYGTAMLCVV-TPKEHLGL 510

Qy  379 PHNYPAHGNVVYS-GTHDNDTTLGWFRTAPEAERAFMRAYLARYGIRCLSEYEVA 433
Db  511 PDKEDVRTGIITYKLAAHAADLAKGW----PGAQLRDNALSKARFEFRWRDQFRLS 562
```

RESULT 5
US-11-075-185-10
; Sequence 10, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
; APPLICANT: REEVES, CHRISTOPHER D
; APPLICANT: JULIEN, BRYAN
; APPLICANT: REID, RALPH
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
; FILE REFERENCE: 010099.03

```
; CURRENT APPLICATION NUMBER: US/11/075,185
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/551,103
; PRIOR FILING DATE: 2004-03-08
; PRIOR APPLICATION NUMBER: US 60/568,290
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 10
; LENGTH: 1766
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-11-075-185-10

Query Match            3.3%;  Score 89;  DB 7;  Length 1766;
Best Local Similarity 18.9%;  Pred. No. 7.2;
Matches 104; Conservative 49; Mismatches 154; Indels 242; Gaps 26;

Qy   90 PPYPTQRVDYGWL----------------YQTRWPLLRRAFAGFRARA------SAQDK 126
Db  895 PTYPWQRESY-WLEAPKSQTPRQHGAEHHYETEWRLAERERPAEPRRGGWLILDDQAERA 953

Qy  127 TRLEAFIEAE--------------------RFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAA 172
Db  954 AALQDYLEARGQTCVRVVAADTYARRGARDY---------------QIDPREPEH 993

Qy  173 LAR----AREELAEEVA---------LYEW------------------IQWL 193
Db  994 FARLLGEQEVVDALADASPSDRCGVVHLWSAHSSPAPTLESIQQAQALGSISALHLVQAL 1053

Qy  194 FYLEWGQ-TKAYAESKGIQIIGDMPIFVAFDSSDVWA--------NPQYFYLEADGNP 242
Db 1054 ARAGWRQPPRLMLVTQEVQAIKNPTVSVA--QAPVWGFGATVALEMPELQCTLLDLDATP 1111

Qy  243 TV------------------------VAGVPRD---------------YPSETGQLWG 261
Db 1112 NIDALGQELLSASDEDRIALRGAERHVARLVPHVPEQRPAPEPLSFKADATYLLTGGLGG 1171

Qy  262 NPL--YRW---------------DVMERDNFAWCI 279
Db 1172 IGLVVLEWMAARGARHFALLGRSGPSASAQPVLDRMREDGAQVRTFSVDVADRERLRTVL 1231

Qy  280 ARIRQSLKQCHLVRIDHFRGPEAYWEVP-------FGRPNAVEGRMVKAPGEKLFAAV 330
Db 1232 AQIQTSMPP---LAGIIHAAGVGDQKMIPDLDGPSLQAIGRPK-VDGSW------NL 1278

Qy  331 RAQLSDAPIIAEDLGVITPEVEALRDGFGFPGMKILQFAFSGEDNAF---LPHNYPAHGN 387
Db 1279 HELTSELPL--DFFVLPSSVSSL---FGSHG----QSSYAA-GNAFLDALSHRRALGL 1327

Qy  388 VVVYSGTHDNDTTLGWFRTAPEAER----APMRAYLARYGIRCLSEYEVAGALIELAFKS 443
Db 1328 PAL-----------SLNWTAWTDVGMATPIIAHTSRYLATQGMGALSREGVAALEQLFRAS 1378

Qy  444 PAKLAIVPL 452
Db 1379 SAQIGVVPL 1387
```

RESULT 6
US-11-055-822-460
; Sequence 460, Application US/11055822
; Publication No. US2005026O707A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
; TITLE OF INVENTION: METABOLIC PATHWAY PROTEINS
; FILE REFERENCE: BGI-121CPCN
; CURRENT APPLICATION NUMBER: US/11/055,822

Query Match   3.2%;   Score 88;   DB 7;   Length 620;
Best Local Similarity   21.4%;   Pred. No. 2.2;
Matches 113;   Conservative   73;   Mismatches 156;   Indels 186;   Gaps   35;

```
Qy      7 FGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQV----LPLGPTSYGDSPYQSF 62
            |||  :||:  | : |||:   |||:      ||| :
Db    160 WGVQFHPESIGGQFGHQII----KNFLNL---ARTYRWQLTEKTIPLSVDS--AAVFETF 210

Qy     63 SAFA--------GNPYLVDPEMLIEK-----------GWL-EQSEAPPYPTQRVDY 99
            :                   |||||                |||
Db    211 FAHSSHAFWLDDAQGTSYLGDASGPLARTKTHNVGEGDFFTWLKEDLAANSVAPGCQGFRL 270

Qy    100 GWLYQTRWPLLRRAFAGFRA----------------RASA--QDKTRLEAFIEAERFWLED 142
            ||| :                              :|          || |  |||||
Db    271 GWVGVYGYEL--KAEAGARAAHTSSLPDAHLIFADRAIAVESDQVRLLALGEQDE-WFEE 327

Qy    143 --------------------YALFMALKTRFDGKPWNEWSPELRDREPAALARAREELAE---EV 184
                                |:|        : |          |: |
Db    328 TIKKLHNLVAPRIPASGHLALQVR-DSK--DEYLDKIR------RAQELITRGESYEI 376

Qy    185 ALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYF--YLEADGNP 242
            |||         :      |  :: |                ::|
Db    377 CLTTKLQ----------------GTTDVAPLAAYLALRG---ANPTAYGAYLQL-GDT 414

Qy    243 TVVAGVPRDYFS--ETGQLWGNPLY----RWDVMERDNFAWCIARIR------QSLKQC 289
            :::: :|        |||          || |  :      :::: :
Db    415 SILSSSPERFITIDSAGYVESKPIKGTRPRGRTAQEDQ--EIIAELRSNPKDRAENLMIV 472

Qy    290 HLVRIDHFRGFEAYWEVPFGRPNAVEGRWVKAPGEKLF-----AAVRAQLSDAPIIAEDL 344
            |||  :    |                :|| :|           :|
Db    473 DLVRNDLARG--------ALPTTVK------TSKLFDVETYATVHQLVS--TVSAEL 513

Qy    345 GVITPEVEALRDGF------GFPGMKILQFAFSGEDNAFLPHNYPAHGNVVVYSGTHDND 398
```

RESULT 7
US-11-055-822-702, Sequence 702, Application US/11055822

Db　　514　GPRSP-IECVRAAFPGGSMTGAPKLRTMEII---DELEAAPRG------IYSG----- 556

Qy　　399　TTLGWFRTAPEAERAFMRAYLARYGIRCL------SEYEVAGALIELA 440

Db　　557　-GLGYFSLDGAVDLSMV-------IRTLVIQNNHVEYGVGGALLAIS 595

RESULT 8
US-11-167-856-20
; Sequence 20, Application US/11167856
; Publication No. US20050268352A1
; GENERAL INFORMATION:
; APPLICANT: Nikolau, Basil J
; APPLICANT: Wurtele, Eve S
; APPLICANT: Oliver, David J
; APPLICANT: Behal, Robert
; APPLICANT: Schnable, Patrick S
; APPLICANT: Ke, Jinshan
; APPLICANT: Johnson, Jerry L
; APPLICANT: Allred, Carolyn C
; APPLICANT: Fatland, Beth
; APPLICANT: Lutziger, Isabelle
; APPLICANT: Wen, Tsui-Jung
; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
; TITLE OF INVENTION: Acetyl CoA Levels in Plants
; FILE REFERENCE: P2194USDIV-2
; CURRENT APPLICATION NUMBER: US/11/167,856
; CURRENT FILING DATE: 2005-06-27
; PRIOR APPLICATION NUMBER: US 10/293,865
; PRIOR FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/344,882
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 60/090,717
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 20
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Arabidopsis Thaliana
US-11-167-856-20

Query Match　　　　3.2%;　　Score 87;　DB 7;　Pred. No. 2.2;
Best Local Similarity　　27.5%;
Matches　39;　Conservative　14;　Mismatches　47;　Indels　42;　Gaps　8;

Qy　　148　ALKTRFDGKPWNEWSPELRDREPAALARAREELAEEVALYEWIQWLFYLEWGQTKAYAES 207

Db　　103　AARTAFDEGPWPKMSAYERSRVLLRPADLVEKHSEELASLE-------TWDNGKPYQQS 154

Qy　　208　KGIQIGDMPIFV-AFPDSSDVWANPQY-FYLEADGN--------PTVVAG--VPRDYFSE 255

Db　　155　---LTAEIPMFARLFRYYAGWADKIHGLTIPADGNYQVHTLHEPIGVAGQIIP------- 204

Qy　　256　TGQLWGNPLYRWDVMERDNFAW 277

Db　　205　----WNFPLLM-------FAW 214

RESULT 9
US-11-012-762-64
; Sequence 64, Application US/11012762
; Publication No. US20050244815A1
; GENERAL INFORMATION:
; APPLICANT: Georgia State University Research Foundation, Inc.
; TITLE OF INVENTION: Compositions and Methods for Viral Resistance Genes
; FILE REFERENCE: GSUI.PCT
; CURRENT APPLICATION NUMBER: US/11/012,762
; CURRENT FILING DATE: 2004-12-15
; PRIOR APPLICATION NUMBER: PCT/US03/19300
; PRIOR FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US 60/390,046

; PRIOR FILING DATE: 2002-06-19
; NUMBER OF SEQ ID NOS: 133
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 64
; LENGTH: 512
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-11-012-762-64

Query Match　　　　3.2%;　　Score 86;　DB 7;　Length 512;
Best Local Similarity　　18.1%;　Pred. No. 2.6;
Matches　79;　Conservative　45;　Mismatches　97;　Indels　216;　Gaps　21;

Qy　　110　LRRAFAGFRARASAQDKTRLRAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDRE 169

Db　　191　LQRSFVKGHR------PTKLKSLLRLVKHWYQQY----------VRDKC 222

Qy　　170　PAALARAREELAEEVALYEW----------IQWLFYLEWG 199

Db　　223　PRANLPPLYAL-ELLTVYAWRAGTQEDSNFRLDEGLATVMELLQDHELLCIYW-------- 274

Qy　　200　QTKAYAESKGI------QIIGDMPIFVAFDSSDVWANPQYFYLEADGNPTVVAGVPRDY 252

Db　　275　-TKYYTLQHPVIERFVRRQLKGERPIIL--------DP-----ADPTHNVAQG----- 313

Qy　　253　FSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQ--CH------LVRIDHFRG 299

Db　　314　------------YRWDIVAQ------RASQCLKQDCCYDDRDAPVPSWTVKRAPYIQV 353

Qy　　300　FEAYWEVPFGRPNAVEGRWVK--APGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDG 357

Db　　354　TVQQW---GHPDLI--LWVNPYEPIKKL---------------KEKIRLS 383

Qy　　358　FGPPGMKILQFAFSGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRA 417

Db　　384　RGYSGLQRLSF-----------------------QBPGGQRQLIRS 406

Qy　　418　Y--LARYGIRCLSE---YEVAGALIELAFKSP----AKLAIVPLQDVLGL---- 458

Db　　407　QCSLAVYGIFCDTQICLLDTISPEIQVFVKNPDGGSHAYAIHPLDFVLSLKQQIEDRQGL 466

Qy　　459　-GPEARMNFPGRLGDNW 474

Db　　467　QSQEQQLEFQGRVLEDW 483

RESULT 10
US-11-038-284-34
; Sequence 34, Application US/11038284
; Publication No. US2005024679A1
; GENERAL INFORMATION:
; APPLICANT: COOKE, DAVID
; APPLICANT: DEBET, MARTINE
; APPLICANT: GIDLEY, MICHAEL, JOHN
; APPLICANT: JOBLING, STEPHEN, ALAN
; APPLICANT: SAFFORD, RICHARD
; APPLICANT: SIDEBOTTOM, CHRISTOPHER, MICHAEL
; APPLICANT: WESTCOTT, ROGER, JOHN
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO PLANT STARCH COMPOSITION
; FILE REFERENCE: 054163-5003-US
; CURRENT APPLICATION NUMBER: US/11/038,284
; CURRENT FILING DATE: 2005-01-21
; PRIOR APPLICATION NUMBER: US/11/056,454
; PRIOR FILING DATE: 2002-01-24
; PRIOR-APPLICATION NUMBER: PCT/GB96/01075
; PRIOR FILING DATE: 1996-05-03
; PRIOR APPLICATION NUMBER: GB 9607409.1
; PRIOR FILING DATE: 1996-04-10
; PRIOR APPLICATION NUMBER: GB 9509229.2
; PRIOR FILING DATE: 1995-05-05
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34

```
;  LENGTH: 695
;  TYPE: PRT
;  ORGANISM: Solanum tuberosum
US-11-038-284-34

Query Match          3.2%;  Score 85.5;  DB 7;  Length 695;
Best Local Similarity 18.3%;  Pred. No. 4.3;
Matches   67;  Conservative   40;  Mismatches  106;  Indels  153;  Gaps  18;

Qy   54 YGDSPYQSFSAFA-----GNP----YLVDP-----EMLIE-----------KGW-L  83
           ||| ::|| ||      |||    :|| ||    ::|: :
Db  218 YGSFGYHVTNFFAVSNRYGNPEDLKYLIDKAHSLGLQVLVDVVHSHASNNVTDGLNGFDI 277

Qy   84 EQSEAPPPYTQRVDYGWLIYQTR------WPLLRRAFAGFRARASAQDKTRLEAFIEAER 137
         :|:                       :||||||||||||||||||| :
Db  278 GQGSQESYFHAGERGYHKLWDSRLFNYANWEVLRFLLSNL------------------R 318

Qy  138 FWLEDYAL--FMALKTRFDG----------------KPNNEWSPELRDREPAALARAREE 179
           ||| |      |||||||||
Db  319 WWLEEYNFDGF------RFDGITSMLYVHHGINMGFTGNYNEYFSEATDVDAVVYLMLANN 373

Qy  180 L----------AEEVALYEWI------------QWLFYL-----EWGQT 201
         |          ||:|:                 :||
Db  374 LIHKIFPDATVIAEDVSGMPGLSRPVSEGGIGFDYRLAMAIPDKWIDYLKNKNDEDWSMK 433

Qy  202 K----------AYAESKGIQIGDMPIFVAFDSSDVWANPQYFYLEADGNPTVVA 246
         :          ||||| |         ::||        :|:
Db  434 EVTSSLTNRRYTEKCIAYAESHDQSIVGDKTIAFLLMDKEMYSGMSCL---TDASPVVDR 490

Qy  247 GVP----------RDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLV 292
         ||            |||||   |||||    :                 :
Db  491 GIALHKMIHFFTMALGGEGYLNFMGNEFGHP--EWIDFPREGNNWSYDKCR---RQWNLA 545

Qy  293 RIDHFR 298
         :| | |
Db  546 DSEHLR 551

RESULT 11
US-11-205-109-14
;  Sequence 14, Application US/11205109
;  Publication No. US2005028764lA1
;  GENERAL INFORMATION:
;  APPLICANT: Farnet, Chris
;  APPLICANT: Zazopoulos, Emmanuel
;  APPLICANT: Staffa, Alfredo
;  TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS
;  FILE REFERENCE: 3002-2US
;  CURRENT APPLICATION NUMBER: US/11/205,109
;  CURRENT FILING DATE: 2005-08-17
;  PRIOR APPLICATION NUMBER: US/09/976,059
;  PRIOR FILING DATE: 2001-10-15
;  PRIOR APPLICATION NUMBER: US 60/239,924
;  PRIOR FILING DATE: 2000-10-13
;  NUMBER OF SEQ ID NOS: 46
;  SOFTWARE: Patentin version 3.0
;  SEQ ID NO 14
;  LENGTH: 6893
;  TYPE: PRT
;  ORGANISM: Actinoplanes sp.
US-11-205-109-14

Query Match          3.1%;  Score 85;  DB 7;  Length 6893;
Best Local Similarity 19.8%;  Pred. No. 1e+02;
Matches  101;  Conservative   44;  Mismatches  153;  Indels  212;  Gaps  26;

Qy   17 PGRWGIGAL----GREAERFLD-WLADAGAWWQVLPLGPTSYGDSPYQSFSAFAGNPYL  71
         |||  |||     ||:|| || ||
Db 1515 PGR-ALGTLDVVAGRERDLLLDGWNA------------------TAVPAQPAL 1548

Qy   72 VDPEMLIEKGWLEQSEAPPPYPTQRVDYGWLYQTRWPLLRRAFAGFRARASAQDKTRLEA 131
         |  |                 ::|:   ||        |
Db 1549 V-PEL-----FTAQAARTP------------------TWPALVTAGAEMSYAELEERSNRLAR 1587
```

```
Qy  132 FIEAERFWLED----------------YALFMALKTRFDGKPWNEWSPEL-RDREPAALARAR 177
         ::                         |||||||||||| || |
Db 1588 WLAGRGVGADDRVALMMRRGPELMVAILAVLKA---GAAYLPVDPDLPRDRVDYLLADAA 1644

Qy  178 EE--LAEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFY 235
                                          |||  :|:|:  |||||
Db 1645 PAFVLAER-----------------------ATAPWVPVAGGIPVLVV-------------- 1669

Qy  236 LEADGNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRID 295
                         :::                 |:||
Db 1670 ----DAPAVAAEVA---AHSGEA-----------VTDRDR----RAALRGGHLAYVI 1704

Qy  296 HFRGPEAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALR 355
                                                          || |
Db 1705 YTSG------STGRPKGV-----------------LITHD-GLANLTLDHGR 1732

Qy  356 DGFGFPGMKILQFAFSGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAER--A 413
         ||| | | ::||:|                  :
Db 1733 FGLG-PGARVAQFASPGFD--------MFVDEWSMALLAGAA---LTFVPPERRLGA 1777

Qy  414 FMRAYLARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDN 473
                       |:|        ||:|               :|  :
Db 1778 DLAAFLAEYG----------VTHATLPPAVVGTIPDGVLPPSFVLDVVGGDV---LPGDLARR 1826

Qy  474 W----------------------------AWRYAEGD 482
         |                           :   |||||
Db 1827 WLRDGRVLFNSYGPTETTVNAATWRAEAGD 1856

RESULT 12
US-10-454-437-102
;  Sequence 102, Application US/10454437
;  Publication No. US2005027715A1
;  GENERAL INFORMATION:
;  APPLICANT: Pompejus, Markus
;  APPLICANT: Kroger, Burkhard
;  APPLICANT: Schroder, Hartwig
;  APPLICANT: Zelder, Oskar
;  APPLICANT: Haberhauer, Gregor
;  TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
;  TITLE OF INVENTION: INVOLVED IN HOMEOSTASIS AND ADAPTATION
;  FILE REFERENCE: BGI-128CPCN
;  CURRENT APPLICATION NUMBER: US/10/454,437
;  CURRENT FILING DATE: 2003-06-13
;  PRIOR APPLICATION NUMBER: US 60/141031
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: DE 19931636.8
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19932125.6
;  PRIOR FILING DATE: 1999-07-09
;  PRIOR APPLICATION NUMBER: DE 19932126.4
;  PRIOR FILING DATE: 1999-07-09
;  PRIOR APPLICATION NUMBER: DE 19932127.2
;  PRIOR FILING DATE: 1999-07-09
;  PRIOR APPLICATION NUMBER: DE 19932128.0
;  PRIOR FILING DATE: 1999-07-09
;  PRIOR APPLICATION NUMBER: DE 19932129.9
;  PRIOR FILING DATE: 1999-07-19
;  PRIOR APPLICATION NUMBER: DE 19932226.0
;  PRIOR FILING DATE: 1999-07-09
;  PRIOR APPLICATION NUMBER: DE 19932920.6
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: DE 19932922.2
;  PRIOR FILING DATE: 1999-07-14
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 442
;  SEQ ID NO 102
;  LENGTH: 355
;  TYPE: PRT
;  ORGANISM: Corynebacterium glutamicum
US-10-454-437-102
```

Query Match   3.1%; Score 84.5; DB 6; Length 355;
Best Local Similarity 21.2%; Pred. No. 2.1;
Matches 68; Conservative 33; Mismatches 87; Indels 133; Gaps 19;

```
Qy  127 TRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAR---------- 177
Db   68 THADYVIEEET--ADDVAAVRAL----------------LYDLAPAASANAKAEYKLSAEE 110
Qy  178 -EELAEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYL 236
Db  111 AEEHGQKVGAADWSFW-------EAKVRARD--------YALDETEL---RNYFPPL 148
Qy  237 EADGNPTVVAGVPRDYFSETGQLWG------NPLYR-------WDVMERDNFAWCIARIR 283
Db  149 ----NQVLRDGV----FFAANRLYGITVEPRPDLRGYAEGVDVWEVLDSDG--------- 191
Qy  284 QSLKQCHLVRIDHFRGFEAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAED 343
Db  192 ---SGIGLILTDY----------YGRPSKRGGAWMSS-----FVDQSELLGTKPVVVNV 232
Qy  344 LGVITPEV-EALRD----------GFGPFGM--KILQFAFSG----------ED 374
Db  233 MGITKPTTGEALLSLDEVTTIFHEFGHGLHGLLSKVRYPSFSGTSVPRDYVEFPSQINEN 292
Qy  375 NAFLP---HNYPAH--GNVV 389
Db  293 WAFDPAVVRNYARHVDTGDII 313
```

RESULT 13
US-10-454-437-100
; Sequence 100, Application US/10454437
; Publication No. US20050277115A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
; TITLE OF INVENTION: INVOLVED IN HOMEOSTASIS AND ADAPTATION
; FILE REFERENCE: BGI-128CPCN
; CURRENT APPLICATION NUMBER: US/10/454,437
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: DE 19931636.8
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19932125.6
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932126.4
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932127.2
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932128.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932129.9
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: DE 19932226.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932920.6
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19932922.2
; PRIOR FILING DATE: 1999-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 442
; SEQ ID NO 100
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-10-454-437-100

Query Match   3.1%; Score 84.5; DB 6; Length 555;

Best Local Similarity 21.2%; Pred. No. 3.9;
Matches 68; Conservative 33; Mismatches 87; Indels 133; Gaps 19;

```
Qy  127 TRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAR---------- 177
Db  268 THADYVIEEET--ADDVAAVRAL----------------LYDLAPAASANAKAEYKLSAEE 310
Qy  178 -EELAEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYL 236
Db  311 AEEHGQKVGAADWSFW-------EAKVRARD--------YALDETEL---RNYFPPL 348
Qy  237 EADGNPTVVAGVPRDYFSETGQLWG------NPLYR-------WDVMERDNFAWCIARIR 283
Db  349 ----NQVLRDGV----FFAANRLYGITVEPRPDLRGYAEGVDVWEVLDSDG--------- 391
Qy  284 QSLKQCHLVRIDHFRGFEAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAED 343
Db  392 ---SGIGLILTDY----------YGRPSKRGGAWMSS-----FVDQSELLGTKPVVVNV 432
Qy  344 LGVITPEV-EALRD----------GFGPFGM--KILQFAFSG----------ED 374
Db  433 MGITKPTTGEALLSLDEVTTIFHEFGHGLHGLLSKVRYPSFSGTSVPRDYVEFPSQINEN 492
Qy  375 NAFLP---HNYPAH--GNVV 389
Db  493 WAFDPAVVRNYARHVDTGDII 513
```

RESULT 14
US-10-858-730-22
; Sequence 22, Application US/10858730
; Publication No. US2005025568A1
; GENERAL INFORMATION:
; APPLICANT: Bailey, Richard B.
; APPLICANT: Blomquist, Paul
; APPLICANT: Doten, Reed
; APPLICANT: Driggers, Edward M.
; APPLICANT: Madden, Kevin T.
; APPLICANT: O'Leary, Jessica
; APPLICANT: O'Toole, George
; APPLICANT: Trueheart, Joshua
; APPLICANT: Walbridge, Michael J.
; APPLICANT: Yorgey, Peter S.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
; TITLE OF INVENTION: PRODUCTION
; FILE REFERENCE: 14184-030001
; CURRENT APPLICATION NUMBER: US/10/858,730
; CURRENT FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/475,000
; PRIOR FILING DATE: 2003-05-30
; PRIOR APPLICATION NUMBER: US 60/551,860
; PRIOR FILING DATE: 2004-03-10
; NUMBER OF SEQ ID NOS: 364
; SEQ ID NO 22
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-858-730-22

Query Match   3.1%; Score 84; DB 6; Length 379;
Best Local Similarity 18.0%; Pred. No. 2.6;
Matches 93; Conservative 48; Mismatches 139; Indels 236; Gaps 23;

```
Qy   22 IGALGREAERFLDWLADAGARWNQVLPLGPTSYGDSPYQSFSAFAGNPYLVDPEMLIEKG 81
Db   22 VGSLQLESGAVIDDVCIAVQRWGKLSPA-----RDNVVVLHALTGDSHITGP------- 69
Qy   82 WLEQSEAPPPYPTQRVDYGW----------LYQTRW-PLLRRAFAGFRARASAQDKTRLE 130
Db   70 ------AGPGHPTP-----GWWDGVAGPGAPIDTTRWCAVATNVLGGCRGSTGPSSLAR-- 117
Qy  131 AFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARRELABEVALYEWI 190
```

```
Db   214 YR-----SRNNRSYKVYGYSPVEQIIMTVNIALKRQVHALEYYTAGSVPDALVGVPETWSA 269
Qy   295 DHFRGFEAYWEVPFGRPNAVEGRWVKAPGE--KLFAAVRA-----------------Q 333
Db   270 DDIRRFQEYWDLLLSGETAQRRKOMRFVPGELSRNFRETKQPPLKDVYDEWLARVVCFAFS 329
Qy   334 LSDAPIIAEDLGVI--TPEVEALRDGFG------------------FPGMKILQFAFSGEDN 375
Db   330 VEPTFVAQVNRSVAETSREQSLSDGMGSLKNWVKALIDDVLARYMDMAAYEFVWKGEES 389

Search completed: January 10, 2006, 14:25:13
Job time :30 secs
```

```
Db   118 -----DGKPWGSRFPLISIRD------------------ 133
Qy   191 QWLFYLEWGCQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADGNPTVVAGVPR 250
Db   134 -----QVQADVAALAALGITEVAA---------VVG--- 155
Qy   251 DYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGFEAYWEVPFGR 310
Db   156 -----GSMGGARALEWVV-------------------GY 170
Qy   311 PNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGF---------- 360
Db   171 PDRVRAGLLLAVG------------ARATADQIGTQTTQIAAIKADPDWQSGDYHETGR 217
Qy   361 ---PGMKILQ----FAFSGE---DNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEA 410
Db   218 APDAGLRLARRFAHLTYRGEIELDTRFANHN---QGN.----------EDPTAG------- 257
Qy   411 ERAFMRAYLARYGIRCLSEYEVAGALIEL--AFKS------------PAKLAIVPLQDVLGL 458
Db   258 GRYAVQSYLEHQGDKLLSRFD-AGSYVILTEALNSHDVGRGRGGVSAALRACPVPVVVG- 315
Qy   459 GPEARMNFPGRLGDNWAWRYAEGDLEPGLAAGLRAL 494
Db   316 GITSDRLYPLRLQQELA------DLLPG-CAGLRVV 344

RESULT 15
US-10-467-657-70
; Sequence 70, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 70
; LENGTH: 715
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-70

Query Match      3.0%;  Score 82;  DB 6;  Length 715;
Best Local Similarity  19.8%;  Pred. No. 9.3;
Matches  83;  Conservative  41;  Mismatches  136;  Indels  160;  Gaps  23;

Qy    64 AFAGNPYLVDPEMLIEKGWLEQSEAPPPYPTQ----RVDYGWLYQT--RWPLLRRA--FA 115
Db    22 AFTGN----------ADGWFDAGEPPAPAAQQAEGRRFDYEPFYNVGHSKPREREAVGFA 71
Qy   116 GFRARASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALAR 175
Db    72 QLRALADNYDVLRL--VIEARKDQME------CLK------WTIQKRDVESTEDDESQR 116
Qy   176 AREELAEEVALYE------WIQW------LFYLE--------WGQTKAYAESKGIQ 211
Db   117 KDRKVDEAVAFFRSPDKEHTWADWLRILLEDLFVIDAPCIYPRKTLGGGLYALEVMDGAT 176
Qy   212 I------IGDMPIFVAFDSSDVWANPQYFYLEADGNPTVVAGVPR-DYFSETGQLWGNPL 264
Db   177 IKRVLDNTGRMPL---------PPDTAYQQ------ILHGMAAVDYTAD------ELI 213
Qy   265 YRWDVMERDNFAW------------CIARIRQSLK-QCHLVR------------I 294
```

```
                    GenCore version 5.1.6
         Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    January 10, 2006, 13:56:47 ; Search time 134 Seconds
                                        (without alignments)
                                        1642.752 Million cell updates/sec

Title:        US-10-618-976-2
Perfect score: 2714
Sequence:    1 MQLQRAFGILLHPTSFPGRW.........DLEPGLAAGLRALARASQRA 501

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5
Searched:    2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters:      2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   A_Geneseq_21:*
             1: geneseqp19808:*
             2: geneseqp19909:*
             3: geneseqp20008:*
             4: geneseqp20018:*
             5: geneseqp20028:*
             6: geneseqp20038a8:*
             7: geneseqp20038b8:*
             8: geneseqp20048:*
             9: geneseqp20058:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2714	100.0	501	4	AAE01153	Aae01153 Thermus r
2	1696.5	62.5	500	8	AAW83330	Aaw83330 Thermus f
3	1696.5	62.5	500	8	ADQ15011	Adq15011 Thermus a
4	1692.5	62.4	500	7	ADC23307	Adc23307 Protein o
5	1367	50.4	496	8	ADU07729	Adu07729 Amino aci
6	1222.5	45.0	505	8	ADQ15026	Adq15026 Synechocy
7	1196	44.1	505	8	ADK47713	Adk47713 Streptoco
8	1194	44.0	505	8	ABU02622	Abu02622 S. pneumo
9	1108.5	40.8	487	8	ADQ15028	Adq15028 Clostridi
10	1100	40.5	485	7	ADC23309	Adc23309 Protein o
11	1076	39.6	458	9	ADR94304	Adr94304 Novel S.
12	1076	39.6	458	9	AEA58174	Aea58174 Streptoco
13	1041	38.4	497	5	ABP28631	Abp28631 Streptoco
14	1035.5	38.2	576	8	ADQ15022	Adq15022 Potato am
15	1020	37.6	498	5	ABP28630	Abp28630 Streptoco
16	1020	37.6	498	8	ADV88490	Adv88490 Streptoco
17	1020	37.6	498	8	ADV81901	Adv81901 Streptoco
18	1020	37.6	498	8	ADV79743	Adv79743 Streptoco
19	894.5	33.0	486	8	ADX94539	Adx94539 Plant ful
20	777.5	28.6	298	7	ADG89854	Adg89854 Streptoco
21	694	25.6	291	7	ADG89855	Adg89855 Potato 4-
22	605.5	22.3	470	9	ABM90697	Abm90697 M. xanthu
23	451	16.6	526	2	AAY34929	Aay34929 C. pneumo
24	451	16.6	526	5	ABB94330	Abb94330 Chlamydia
25	451	16.6	526	8	ADQ15030	Adq15030 Chlamydia
26	430.5	15.9	527	5	ABB94365	Abb94365 Chlamydia
27	383	14.1	683	8	ABU03132	Abu03132 Alpha amy
28	383	14.1	683	8	ADU07478	Adu07478 Amino aci
29	361	13.3	388	2	AAY37126	Aay37126 Protein i
30	356.5	13.1	698	5	AAM49553	Aam49553 Actinopla
31	333	12.3	183	5	AAW83189	Aaw83189 Corn 4-al
32	333	12.3	727	6	ABM65372	Abm65372 Propionib
33	328	12.1	183	4	ADG89857	Adg89857 Corn 4-al
34	323.5	11.9	204	8	ADX92724	Adx92724 Plant ful
35	313.5	11.6	684	8	ADS14923	Ads14923 Pseudomon
36	301	11.1	696	7	ABM70485	Abm70485 Photorhab
37	299	11.0	681	7	ABO75883	Abo75883 Pseudomon
38	289	10.6	694	8	ADQ15024	Adq15024 Escherich
39	278	10.2	743	7	ABO64556	Abo64556 Klebsiell
40	273.5	10.1	496	8	AAW83188	Aaw83188 Corn 4-al
41	273.5	10.1	496	7	ADG89851	Adg89851 Corn 4-al
42	272	10.0	153	4	AAU03658	Aau03658 Group B S
43	271	10.0	584	4	AAU44938	Aau44938 Propionib
44	271	10.0	584	6	ABM41457	Abm41457 Propionib
45	264	9.7	721	5	ABP66204	Abp66204 Bifidobac

ALIGNMENTS

```
RESULT 1
AAE01153
ID   AAE01153 standard; protein; 501 AA.
XX
AC   AAE01153;
XX
DT   06-AUG-2003  (revised)
DT   17-JUL-2001  (first entry)
XX
DE   Thermus rubens glucanotransferase protein.
XX
KW   Glucanotransferase; starch; Japanese dessert; noodles; wheat product;
KW   sea food; processed food; baby; animal; drink; gyoza skin; shumai skin;
KW   nutrient supplemental food; snack; detergent composition.
XX
OS   Thermus sp.
XX
XX
PN   WO200129195-A1.
XX
PD   26-APR-2001.
XX
PF   16-OCT-2000; 2000WO-DK000560.
XX
PR   20-OCT-1999; 99DK-00001501.
PR   15-NOV-1999; 99DK-00001641.
XX
PA   (NOVO ) NOVOZYMES AS.
XX
PI   Fukuyama S;
XX
DR   WPI; 2001-290912/30.
DR   N-PSDB; AAD04867.
XX
PT   Novel polypeptides having glucanotransferase activity for producing food
PT   and in detergent or cleaning compositions for removing starch stains,
PT   especially amylose stains.
XX
PS   Claim 1; Page 52-54; 57pp; English.
XX
CC   The present sequence is Thermus rubens ATCC 31556 glucanotransferase.
CC   Glucanotransferase catalyses the cyclisation of starch by acting on
CC   starch in food to produce a cyclic glucan. Glucanotransferase is useful
CC   for producing food such as Japanese desserts, snacks, wheat products,
CC   noodles, gyoza skins, shumai skins, processed sea foods, frozen or
CC   refrigerated processed foods, weaning foods, baby foods, pet foods,
CC   animal feeds, drinks, sports foods and nutrient supplemental foods. A
CC   cleaning or detergent composition comprising glucanotransferase is useful
```

```
CC   for removal of starch stains, in particular for removal of amylose
CC   stains, from a hard surface or from laundry, by contacting the amylose
CC   stain-containing hard surface or the amylose stain-containing laundry
CC   with glucanotransferase or its composition. Nucleic acid constructs,
CC   vectors and host cells comprising glucanotransferase nucleic acid
CC   sequence are also provided. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ   Sequence 501 AA;

Query Match             100.0%;  Score 2714;  DB 4;  Length 501;
Best Local Similarity   100.0%;  Pred. No. 1.7e-255;
Matches  501;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          1 MQLQRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQ 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MQLQRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQ 60

Qy         61 SFSAFAGNPYLVDPEMLIEKGWLEQSEAPPPYPTQRVDYGWLYQTRWPLLRRAFAGFRAR 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SFSAFAGNPYLVDPEMLIEKGWLEQSEAPPPYPTQRVDYGWLYQTRWPLLRRAFAGFRAR 120

Qy        121 ASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAREEL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAREEL 180

Qy        181 AEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADG 240

Qy        241 NPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGF 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 NPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGF 300

Qy        301 EAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGF 360

Qy        361 PGMKILQFAFSGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAYLA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 PGMKILQFAFSGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAYLA 420

Qy        421 RYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWRYAE 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 RYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWRYAE 480

Qy        481 GDLEPGLAAGLRALAEASQRA 501
              |||||||||||||||||||||
Db        481 GDLEPGLAAGLRALAEASQRA 501

RESULT 2
AAW83330
ID   AAW83330 standard; protein; 500 AA.
XX
AC   AAW83330;
XX
DT   17-OCT-2003  (revised)
DT   01-MAR-1999  (first entry)
XX
DE   Thermus flavus amylomaltase.
XX
KW   Thermus flavus; amylomaltase; heat resistant; cyclic glucan;
KW   intramolecular transglycosylation; alpha-glucan; food.
XX
OS   Thermus thermophilus.
XX
PN   EP884384-A2.
XX
PD   16-DEC-1998.
XX
PF   13-MAY-1998; 98EP-00250162.
XX
PR   13-MAY-1997; 97JP-00122635.
```

```
PR   07-MAY-1998; 98JP-00125121.
XX
PA   (EZAK ) EZAKI GLICO CO LTD.
XX
PI   Terada Y, Fujii K, Yanase M, Takata H, Takaha T, Okada S;
XX
DR   WPI; 1999-026580/03.
DR   N-PSDB; AAV72539.
XX
PT   New amylo-maltase catalyses intra-molecular trans-glycosylation of alpha-
PT   glucans - used as additive to e.g. rice products, snacks, wheat products,
PT   noodles, processed seafood, frozen or refrigerated foods, baby foods or
PT   drinks.
XX
PS   Claim 3; Page 17-18; 32pp; English.
XX
CC   The present sequence is an amylomaltase isolated from Thermus flavus.
CC   Amylomaltase catalyses intramolecular transglycosylation of alpha-glucans
CC   to generate cyclic glucans, has no hydrolase activity, has an optimum
CC   temperature of 65-70 degrees Celsius, remains active at 60 degrees
CC   Celsius for at least 10 minutes, is inactivated at 100 degrees Celsius
CC   within 15 minutes and has an optimum pH of 5.5. Amylomaltase can be used
CC   to produce a cyclic glucan by cyclising an alpha-glucan and collecting
CC   and purifying the cyclic glucan (especially where the cyclic glucan
CC   comprises a cyclic alpha-1,4-glucan, a branched cyclic glucan and/or a
CC   branching enzyme is also used in the cyclisation step). The amylomaltase
CC   is used to produce food by adding it to a food material before or
CC   immediately after cooking so that the amylomaltase acts on starch in the
CC   food material to produce a cyclic glucan (especially where the food is
CC   selected from rice products, Japanese desserts, snacks, wheat products,
CC   noodles, gyoza skins, shumai skins, processed seafoods, frozen or
CC   refrigerated processed foods, weaning foods, baby foods, pet foods,
CC   animal feeds, drinks, sports foods and nutritional supplements). (Updated
CC   on 17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 500 AA;

Query Match             62.5%;  Score 1696.5;  DB 2;  Length 500;
Best Local Similarity   62.9%;  Pred. No. 3.6e-156;
Matches  315;  Conservative  65;  Mismatches 118;  Indels   3;  Gaps   2;

Qy          1 MQLQRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQ 60
              | |||||| |||||| :| : ||||| | ||| :||: :::|::||||||||| |||||||
Db          1 MELPRAFGLLLHPTSLPGPYGVGVLGREARDFLRFLKEAGGRYWQVLPLGPTYGVGDSPYQ 60

Qy         61 SFSAFAGNPYLVDPEMLIEKGWLEQSEAPPPYPTQRVDYGWLYQTRWPLLRRAFAGFRAR 120
              ||||||||||| |  :  : | : :    |  ||:|||||:||: : | || :||:||::
Db         61 SFSAFAGNPYLIDLRPLAERGYVRLED--PGFPQGRVDYGLLYAWKMPALKEAFRGFKEK 118

Qy        121 ASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAREEL 180
              || ::  : :||  ||  ||||||||||| :   ::|| :|  ::| | | ||:| :| |
Db        119 ASPEERRAFAAFREREAWWLEDYALFMALKGAHGGLPWNRWPLPLRKREEKALREAKSAL 178

Qy        181 AEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADG 240
              |||| ::|: |||| ::||| |||| |:| :||||||||||| |:|: || |:| ||:||
Db        179 AEEVAFHAFTQWLFFRQWGALKAEAERALGIRIIGDMPIFVAEDSAEYWAHPEWFHLDEG 238

Qy        241 NPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGF 300
              ||||||||| ||||||| :|||||||||||:|| :::| | |:: |  :|||||||||||
Db        239 RPTVVAGVPPDYFSETGQRWGNPLYRWDVLEREGFSFWIRRLEKALELFHLVRIDHFRGF 298

Qy        301 EAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGF 360
              |||||:|  :  ||||||||||||||| ::: : : | :::||||||||||||||| ||:
Db        299 EAYWEIPASCPTAVEGRWVKAPGEKLFQKIQEVFGEVFPVLAEDLGVITPEVEALRDRFGL 358

Qy        361 PGMKILQFAF-SGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAYL 419
              ||||:|||| . :: :|||||||||||.|||| |||||||||||:|||||: ||| |||
Db        359 PGMKVLQFAPDDGMENPFLPHNYPAHGRVVVYTGTHDNDTTLGWYRTATPHEKAFMARYL 418

Qy        420 ARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWRYA 479
              | :|| :: ||  | | : : ::  |:||: | ||||.||.||||| |||..||||||
Db        419 ADWGITFREEEEVPWALMHLGMKSVARLAVYPVQDVLALGSEARMNYPGRPSGNWAWRLL 478
```

```
Qy   480 EGDLEPGLAAGLRALAEASQR 500
         |::| | | |||:|||::|
Db   479 PGELSPEHGARLRAMAEATER 499


RESULT 3
ADQ15011
ID   ADQ15011 standard; protein; 500 AA.
XX
AC   ADQ15011;
XX
DT   23-SEP-2004 (first entry)
XX
DE   Thermus aquaticus amylo maltase polypeptide.
XX
KW   Amylo maltase; transgenic; foodstuff; food additive; modifier;
KW   saccharide; (alpha)-1,4-glucan; cooked rice; Japanese confectionery;
KW   snack; noodle; Chinese dumpling; shao-mai; fishery kneading;
KW   refrigerated; baby food; animal feed; drink; food supplement;
KW   cyclic glucan; Thermus aquaticus; enzyme.
XX
OS   Thermus aquaticus.
XX
PN   JP2004187674-A.
XX
PD   08-JUL-2004.
XX
PF   12-SEP-2003; 2003JP-00322319.
XX
PR   26-NOV-2002; 2002JP-00342966.
XX
PA   (EZAK ) EZAKI GLICO CO LTD.
PA   (NIDE ) NEC CORP.
XX
DR   WPI; 2004-528719/51.
DR   N-PSDB; ADQ15010.
XX
PT   New mutated amylo maltase polypeptide having increased enzyme activity
PT   and reduced hydrolyzing activity, useful for manufacturing foodstuffs
PT   such as Japanese confectionery, noodles, baby foods, and food additives.
XX
PS   Claim 5; SEQ ID NO 2; 51pp; Japanese.
XX
CC   The invention relates to a novel amylo maltase polypeptide comprising the
CC   amino acid sequence of a wild-type amylo maltase with a substitution,
CC   addition or deletion at a position, where the amino acid residue
CC   interacts with acarbose of amino acids other than the wild-type amylo
CC   maltase amino acid sequence. The invention further comprises: a nucleic
CC   acid molecule containing a nucleic acid sequence encoding the novel amylo
CC   maltase polypeptide; a vector containing the nucleic acid sequence; a
CC   cell containing the nucleic acid sequence; biological tissue containing
CC   the nucleic acid sequence; a transgenic organism containing the nucleic
CC   acid sequence; a foodstuff, a food additive or a modifier of foodstuff
CC   containing the nucleic acid sequence; a computer readable recording
CC   medium which contains the information of the nucleic acid sequence
CC   encoding the amino acid sequence of the novel amylo maltase polypeptide;
CC   and a cyclic glucan obtained by reacting the novel amylo maltase
CC   polypeptide on saccharides which have a linear structure of (alpha)-1,4-
CC   glucan. The novel amylo maltase polypeptide is useful for manufacturing
CC   foodstuffs such as cooked rice, Japanese confectionery, snacks, bakeries,
CC   noodles, Chinese dumpling, shao-mai, fishery kneading goods,
CC   refrigerating foods, baby foods, pet foods, feed for animals, drinks, and
CC   food supplements, where the method involves adding the novel amylo
CC   maltase polypeptide to the foodstuff before or immediately after heat
CC   processing of the raw material, where the novel amylo maltase polypeptide
CC   generates a cyclic glucan from the starch of the foodstuff. The novel
CC   amylo maltase polypeptide is useful for manufacturing glucan which has a
CC   cyclic structure of (alpha)-1,4-glycoside linkage, foodstuff materials
CC   and food additives, modifier of foodstuffs, a drink or eatable
CC   composition, infusion solution or composition for adhesion, where the
CC   method involves reacting the saccharide containing the linear structure
CC   of (alpha)-1,4-glycoside linkage or its derivatives with the Thermus aquaticus amylo
CC   maltase polypeptide. This sequence represents the Thermus aquaticus amylo
```

```
CC   maltase polypeptide of the invention.
XX
SQ   Sequence 500 AA;

Query Match             62.5%;  Score 1696.5;  DB 8;  Length 500;
Best Local Similarity   62.9%;  Pred. No. 3.6e-156;
Matches  315;  Conservative  65;  Mismatches  118;  Indels  3;  Gaps  2;

Qy     1 MQLQRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQ 60
           ||:|||||||||||:|||| |||   ||: :|||||:|:|:|| :||||||||||||||
Db     1 MELPRAFGLLLHPTSLPGPYGVGVLGREARDFLRPLKEAGGRYWQVLPLGPTGYGDSPYQ 60

Qy    61 SFSAFAGNPYLVDPEMLIEKGWLEQSEAPPPYPTQRVDYGWLYQTRWPLLRRAFAGFRAR 120
           ||||||||||||||: |||:: :   |:||||  | ||||||||::|||:||: :|:::|
Db    61 SFSAFAGNPYLIDLRPLAERGYVRLED--PGFPQGRVDYGLLJYAWKWPALKEAFRGFKEK 118

Qy   121 ASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAREEL 180
           : :||  :||::|| :|:||| ||:|||:| :||:|:|| ||||:|: :||:|| :| ::|
Db   119 ASPEERRAFAALFREREAWWLEDYALFMALKGAHGGLPWNRWPLPLRKRKEEKALREAKSAL 178

Qy   181 AEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAPDSSDVWANPQYFYLEADG 240
           ||||| |||:  :|:||||||:||:::||| |||||||||| ||:||:||::|| |:: ::
Db   179 AEEVAFHAFTQWLFFPRQWGALKAEEALGIRIIGDMPIFVAEDSAEVWAHPEWFHLDEEG 238

Qy   241 NPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGF 300
           ||||||||||||||||||| |||||||| ||::::::| : ||||:||: |||||||||
Db   239 RPTVVAGVPPDYFSETGQRWGNPLYRWDVLEREGFSFWIRRLEKALELFHLVRIDHFRGF 298

Qy   301 EAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGF 360
           |||| |||  |||||||||||||||| |  ::   |||| |||||||||||||||||:|
Db   299 EAYWEIPASCPTAVEGRWVKAPGEKLFQKIQEVFGEVPVLAEDLGVITPEVEALRDRFGL 358

Qy   361 PGMKILQFAF-SGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAYL 419
           ||||:|||||  || ||||||||||| |||:|||||||||||||||:||  || :||||||
Db   359 PGMKVLQFAFDDGMENPFLPHNYPAHGRVVVYTGTHDNDTTLGWYRTATPHEKAFMARYL 418

Qy   420 ARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWRYA 479
           ||| ::|| | |:: :| |  | | |  |:|  |||:| ||: |:|||||| ||: :|||
Db   419 ADWGITFREEEVPWALMHLGMKSVARLAVYPVQDVLALGSEARMNYPGRFGSGNWAWRLL 478

Qy   480 EGDLEPGLAAGLRALAEASQR 500
           |::| | | |||:|||::|
Db   479 PGELSPEHGARLRAMAEATER 499
```

```
RESULT 4
ADC23307
ID   ADC23307 standard; protein; 500 AA.
XX
AC   ADC23307;
XX
DT   18-DEC-2003 (first entry)
XX
DE   Protein of Thermus thermophilus AMase.
XX
KW   recombinant; alpha-glucanotransferase activity; hydrolysing; hydrolytic;
KW   alphalpha/betabeta; barrel structure; transferase; bakery product; bread;
KW   amylomaltase; branching enzyme; thermostable transferase;
KW   thermophilic microorganism; EC 2.4.1.25; EC 2.4.1.18;
KW   starch retrogradation; amylopectin; alpha-amylase; AMase.
XX
OS   Thermus thermophilus.
XX
PN   WO2003002728-A2.
XX
PD   09-JAN-2003.
XX
PF   01-JUL-2002; 2002WO-NL000427.
XX
PR   29-JUN-2001; 2001NL-01018426.
XX
PA   (NEDE ) NEDERLANDSE ORG TOEGEPAST.
XX
```

```
Db     419  ADWGITFREEEVPWALMHLGMKSVARLAVYPVQDVLALGSEARMNYPGRPSGNWAWRLL  478

Qy     480  EGDLEPGLAAGLRALAEEASQR  500
                ||||||| |  ||| ||::|
Db     479  PGELSPEHGARLRAMAEATER  499
```

RESULT 5
ADU07729
ID   ADU07729 standard; protein; 496 AA.
XX
AC   ADU07729;
XX
DT   13-JAN-2005 (first entry)
XX
DE   Amino acid sequence for amylase #171.
XX
KW   Amylase; glucoamylase; 1,4-alpha-D-glucan glycohydrolase; alpha-amylase;
KW   exoamylase; beta-amylase; glucosidic bond hydrolysis; starch; sugar;
KW   glucose; maltodextrin; thermostable; enzyme.
XX
OS   Unidentified.
XX
PN   WO2004091544-A2.
XX
PD   28-OCT-2004.
XX
PF   08-MAR-2004; 2004WO-US007096.
XX
PR   06-MAR-2003; 2003US-0038530S.
PR   28-MAR-2003; 2003US-0459014P.
XX
PA   (DIVE-) DIVERSA CORP.
XX
PI   Callen W,   Richardson T,   Frey G,   Gray K,   Kerovuo JS,   Slupska M;
PI   Barton N,   O'donoghue E,   Miller C;
XX
DR   WPI; 2004-775532/76.
DR   N-PSDB; ADU07728.
XX
PT   Novel recombinant amylases, useful for hydrolysis of starch into sugars,
PT   in detergent composition, for producing high-maltose or high-glucose
PT   syrup, in brewing or alcohol production, bio-bleaching composition of
PT   paper or pulp product.
XX
PS   Claim 55; SEQ ID NO 439; 436pp; English.
XX
CC   The invention relates to the isolation of polypeptides with amylase
CC   activity, and polynucleotide sequences encoding the polypeptides. The
CC   amylase activity comprises a glucoamylase activity, 1,4-alpha-D-glucan
CC   glycohydrolase activity, alpha-amylase activity, exoamylase activity or
CC   beta-amylase activity. The amylase activity comprises hydrolyzing
CC   glucosidic bonds in a starch to produce sugars such as glucose and
CC   maltodextrins. The amylase activity comprises cleaving a maltose or a D-
CC   glucose unit from the non-reducing end of the starch. The amylase
CC   activity is thermostable and the polypeptides of the invention are active
CC   over a wide range of temperatures. The polypeptide sequences are useful
CC   for making anti-amylase antibodies, for identifying a modulator of an
CC   amylase activity, or identifying an amylase substrate. The polynucleotide
CC   sequences are useful for producing recombinant polypeptides, and for
CC   modifying codons in a nucleic acid encoding a polypeptide with an amylase
CC   activity to increase or decrease its expression in a host cell. The
CC   polypeptide and polynucleotide sequences are useful for making or
CC   modifying a small molecule. Primers to the polynucleotides are useful for
CC   an amylase enzyme. Primers to the polynucleotides are useful for
CC   amplifying them. This sequence represents a polypeptide sequence of the
CC   invention.
XX
SQ   Sequence 496 AA;

```
Query Match     50.4%;  Score 1367;  DB 8;  Length 496;
Best Local Similarity   51.9%;  Pred. No. 5.3e-124;
Matches 260;  Conservative 77;  Mismatches 150;  Indels 14;  Gaps 5;
```

```
PI     Van Der Maarel MJEC,  Dijkhuizen L,  Binnema DJ,  Van Der Veen BA;
PI     Vos A;
XX
DR     WPI; 2003-201501/19.
DR     N-PSDB; ADC23306.
XX
PT     Novel isolated or recombinant nucleic acid encoding a polypeptide with
PT     alpha glucanotransferase activity, useful for reducing the retrogradation
PT     of starch, amylopectin, and for hydrolysing starch.
XX
PS     Disclosure; Page 79; 92pp; English.
XX
CC     The invention relates to a novel isolated or recombinant nucleic acid
CC     derived from a nucleic acid encoding a polypeptide essentially having
CC     alpha-glucanotransferase activity but having essentially no hydrolysing
CC     activity, and the recombinant nucleic acid encoding a polypeptide with
CC     hydrolytic activity. The invention further comprises a vector comprising
CC     the recombinant nucleic acid; a host cell comprising the vector or the
CC     recombinant nucleic acid. The invention also provides a polypeptide or
CC     its fragment essentially having alpha-glucanotransferase activity and
CC     with or without hydrolysing activity, by providing a nucleic acid
CC     encoding such a transferase with a mutation leading to an alteration or
CC     loss of a codon originally encoding a hydrophobic amino acid located in
CC     or around an acceptor, a donor or a catalytic site extending from a
CC     (alphalpha/betabeta) barrel structure of the transferase. The invention
CC     also relates to a bakery ingredient comprising the polypeptide with alpha
CC     -glucanotransferase activity and a bakery product such as bread,
CC     comprising the polypeptide. The transferase encoded by the recombinant
CC     nucleic acid comprises amylomaltase, a branching enzyme or thermostable
CC     transferase. The transferase is derived from a thermophilic microorganism
CC     e.g. Thermus thermophilus, T.aquaticus or Aquifex aeolicus. The
CC     transferase with enzyme commission (EC) number 2.4.1.25 or 2.4.1.18 is
CC     useful in reducing retrogradation of starch and amylopectin, preferably
CC     in reducing long-term retrogradation of amylopectin. This transferase is
CC     also useful for hydrolysing starch. This sequence represents a Thermus
CC     thermophilus AMase protein of the invention.
XX
SQ     Sequence 500 AA;
```

```
Query Match     62.4%;  Score 1692.5;  DB 7;  Length 500;
Best Local Similarity   62.7%;  Pred. No. 8.9e-156;
Matches 314;  Conservative 66;  Mismatches 118;  Indels 3;  Gaps 2;
```

```
Qy      1  MQLQRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQ   60
           | ::|||||||| |||||||  | ::| | |::|||| |:||||:|||||||||||||||
Db      1  MELPRAFGLLLHPTSLPGPYGVGVGLGQEEARDFLRFLKEAGGRYWQVLPLGPTGYGDSPYQ   60

Qy     61  SFSAFAGNPYLVDPEMLIEKGWLEQSEAPPPYPTQRVDYGWLYQTRWPLLRRAFAGPRAR  120
           ||||||||||| : |  |:|||||: ::|:
Db     61  SFSAFAGNPYLIDLRPLAERGYVRLED--PGPPQGRVDYGLLYAWKWPALKEAFRGFKEK  118

Qy    121  ASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPALARAREEL  180
           |  :::  |||:||||:|||||||||||  | ||
Db    119  ASPEEREAFAAFRERAWWLEDYALFWALKGAHGGLPWNRWPLPLRKREEKALREAKSAL  178

Qy    181  AEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAPDSSDVWANPQYFYLEADG  240
           |||| ||||||||| :|| | : ||::|:|: : ::|:||  : :|
Db    179  AEEVAFHAFTQWLFFRQWGALKAEABALGIRIIGDMPIFVAEDSAEVWAHPEWFHLDEEG  238

Qy    241  NPTVVAGVPRDYFSETGQLWGNPLYRWNDVMERDNFAWCIARIRQSLKQCHLVRIDHPRGF  300
           || ||||||| ||||||| |||||||| :||| || || | || ||:||| |||||||||
Db    239  RPTVVAGVPPDYFSETGRWGNPLYRWNDVLERRGFSFWIRRLEKALELFHLVRIDHPRGF  298

Qy    301  EAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGF  360
           ||||||||||||| |||||||||||||  ::||||:|||||||||||||||||:|:|:||
Db    299  EAYWEIPASCPTAVEGRWVKAPGEKLFQKIQSEVFGEVPVLAEDLGVITPEVEALRDRFGL  358

Qy    361  PGMKILQFAF-SGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAYL  419
           ||||:||||  :| |||||||||||||  |: :|||||||||||||| | | ||||||||
Db    359  PGMKVLQFAFDDGMENPLPHNYPAHGRVVTTGTHDNDTTLGWYRTATPHEKAFMARYL  418

Qy    420  ARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWRYA  479
           |::|| | || ||  |:|:| |:: |:|| |||: ||||| |||||| |||:|||||||
```

```
Qy     1 MQLQRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQ 60
         :|:|||||||||||||:||:|: : : ||||::||||||||||||||||||
Db     1 MKFKRAAGIILLHPTSLPGPDGIGDLGPEAYRWINFLAESGCKLWQVLPLGPTGYGDSPYQ 60

Qy    61 SFSAFAGNPYLVDPEMLIEKGWLEQSEAP--PPYPTQRVDYGWLYQTRWPLLRRAFAGFR 118
         ||||||||||:||: ||: : :|| |||||||||||:|:|
Db    61 CFSAFAGNPYLVSPALLLENGLLRRSDLTDRPEFNPIRIDYGEAITWKLKILDRAYENFK 120

Qy   119 ARASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARARE 178
         :: ||| |||:|||||| ||:|:: : || ||||:|||:
Db   121 RLNNPELKYEYQQFTENETDWLNDFALFMAIKEANGGVSWDNWPKELRSRQPDAIEKFKQ 180

Qy   179 ELAEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAPDSSDVWANPQYFYLEA 238
         |||::||||||||| ||||:|: :|||| ||:|||||| ::: ::
Db   181 TEADRIQRHAFRQFLFFRQWLDLKAYANQKNIQIIGDPIFVAYDSADAWNPELFYLNE 240

Qy   239 DGNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFR 298
         :||||||||||||||||||||||||||:|:: :||| ||: :|||:|||||||||||||
Db   241 EGKPTVVAGVPPDYFSPTGQLWGNPLYKWEVHRQQNFRWWIRRMQATLRLVDIVRLDHFR 300

Qy   299 GFEAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGF 358
         ||:|||||||:|| ||:||:||||:|||| ||||:||: :|||||||||||||||||:|
Db   301 GFAGYWEVPFGMPTAEIGRWVKGPGKELFNAIRDALGDLPIIAEDLGVITPDVIELRDSF 360

Qy   359 GFPGMKILQFAFSGED-NAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEARRAFMRA 417
         ||||||||:|:|:|: : : ||:||| || : :||:||:|:||| ||::||:|:|||
Db   361 NLPGMKVFQFAFTTDPLDPFLPHNYET--NCVAYTGTHDNDTTALGWYQSAPEKERDFIRR 418

Qy   418 YLARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWR 477
         |||| :|: |:||||||||| |:|:||||||: :||||||||||||
Db   419 YLARSG-----EDISWDMIRAVWSSVAVFALAPMQDILSLGNEARMNLPGRASGNWCWR 472

Qy   478 YAEGDLEPGLAAGLRALAEAS 498
         ||||| | :||||||
Db   473 LHPEAFNPVI--LKRLKEVN 490
```

```
Query Match             45.0%;  Score 1222.5;  DB 8;  Length 505;
Best Local Similarity   48.6%;
Matches  242;  Conservative   76;  Mismatches 169;  Indels   11;  Gaps    6;

Qy     4 QRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQSFS 63
         :|:||||||||||||||||:||:|: : : ||||::||||||||||||||||
Db     4 KRCSGILLHPTSLPSRFGIGDLGDGAFQFIDFLADADQSWQILPLGPTGFGNSPYLCYS 63

Qy    64 AFAGNPYLVDPEMLIEKGWLEQS--EAPPPYPTQRVDYGWLYQTRWPLLRRAFAGFRARA 121
         ||||||||: ||: : :|| ||||||||||||:|:|
Db    64 ALAINPWLISLDRLAEEGFLPPSLLDQAPPFTNPRVDYDQAIAYKSQVLKQAFAQFRTNI 123

Qy   122 SAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAREELA 181
         ||| |||:|||||| ||:|:: : || ||||:|||:
Db   124 ELAIEQEFAEFCQAQSDWLADYALFMAIKEAHNGAGWHQWDKDIAWREPEALKIWGDRLK 183

Qy   182 EEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAPDSSDVWANPQYFYLEAD-G 240
         |||::||||||||| ||||:|: :|||| ||:|||||| ::: ::
Db   184 TEVLYHQPLQFLGFRQWQEVKAYANQRHIAIFGDLPIYVAHDSADVWANPENFCLDPETG 243

Qy   241 NPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGF 300
         :||||||||||||||||||||||||||:|:: :||| ||: :|||:|||||||||||||
Db   244 EAAMMAGVPPDYFSATGQLWGNPVTDWETLKATGFAWWIKRFKANLQYLDIVRIDHFRGF 303

Qy   301 EAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSD-APIIAEDLGVITPEVEALRDGFG 359
         :||| |||:|| ||:||:||||:|||| ||||:||: :||||||||||||||||:|
Db   304 ESYWGVPQGEKTAERGEWYPAPGKFFQALGKALGDNLPIVAEDLGVITPEVEALRDEFN 363

Qy   360 FPGMKILQFAF-SGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEARAFMRAY 418
         ||||||:|:|: : : ||:||| || : :||:||:|:||| ||::||:|:||||
Db   364 FPGMKVLHFAFPDSDRGNPFLPFNY-SNGNAVVYTGTHDNDTTVGWFQERSEDDQQKVINY 422

Qy   419 LARYGIRCLSEYEVAGALIBLAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWRY 478
         |||| :|: ||:||||||||||||||| |:: ||||||||||||
Db   423 LG-----CVCNEGIHWSLIRLASSSVAALAIFPLQDILGLGSDCRMNLPGTAAGNWGWRY 477
```

```
Qy  479 AEGDLEPGLAAGLRALAE 496
          :|| :| :|
Db  478 HPDQLNDWLSGHLSFITE 495
```

```
RESULT 7
ADK47713
ID   ADK47713 standard; protein; 505 AA.
XX
AC   ADK47713;
XX
DT   20-MAY-2004  (first entry)
XX
DE   Streptococcus pneumoniae protein, Seq ID No 4228.
XX
KW   Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.
XX
OS   Streptococcus pneumoniae.
XX
PN   US6699703-B1.
XX
PD   02-MAR-2004.
XX
PF   26-MAY-2000; 2000US-00583110.
XX
PR   02-JUL-1997; 97US-0051553P.
PR   12-MAY-1998; 98US-0085131P.
PR   30-JUN-1998; 98US-00107433.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CE;
XX
DR   WPI; 2004-212399/20.
DR   N-PSDB; ADK45552.
XX
PT   New nucleic acid molecules and polypeptides useful for diagnosing,
PT   preventing and treating pathological conditions resulting from bacterial
PT   infection, e.g. Streptococcus pneumoniae infection, and in drug
PT   screening.
XX
PS   Disclosure; SEQ ID NO 4228; 301pp; English.
XX
CC   The invention relates to isolated Streptococcus pneumoniae nucleic acids
CC   and polypeptides. The nucleic acids and proteins are useful for
CC   diagnosing, preventing and treating pathological conditions resulting
CC   from bacterial infection, such as S. pneumoniae infection. These may also
CC   be used for drug screening procedures. The present sequence represents a
CC   Streptococcus pneumoniae polypeptide of the invention. Note: The sequence
CC   data for this patent did not appear in the printed specification but was
CC   obtained in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 505 AA;

Query Match      44.1%;  Score 1196; DB 8; Length 505;
Best Local Similarity   48.1%;  Pred. No. 2.7e-107;
Matches  242;  Conservative  77;  Mismatches 168;  Indels  16;  Gaps   9;

Qy    4 QRAFGILLHPTSFPGRWGIGALGRRAERFLDWLADAGARWWQVLPLGPTSYGDSPYQSFS 63
          |:||:||.||||||::|||| ||||  :|||||||||||:||||||||||||
Db    3 KRQSGVLMHISSLPGAYGIGSFGQSAYDFVDFLVRTKQRYWQILPLGATSYGDSPYQSFS 62

Qy   64 AFAGNPYLVDPEMLIEKGWLEQSEAPP---PYPTQRVDYGWGWLYQTRWPLLRRAFAGFRAR 120
          :||||:   :.|  :||||||  :        ||||  || :|| ||||||:|| ||| |
Db   63 AFAGNTHFIDLDILVEQGLLEASDLEGVDFGSDASEVDYAKIYYARRPLLEKAVKRFFEV 122

Qy  121 ASAQDKTRLEAFIEABRPWLEDYALFMALKTRFDGKPWNEW-SPELRDREPAALARAREE 179
           :||||  :||| |::  ::  ::.:||  :: :||:| |:||
Db  123 GDVKD---FEKFAQDNQSWLELFAEYMAIKEHFDNLAWTEWPDADARARKASALESYREQ 179

Qy  180 LAEEVALYEWIQWLFYLEWCGTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEAD 239
          |::::  ||| ::|||||||:|||||  |||||||||||  :||||||||||:|||||
```

```
Db  180 LADKLVYHRVTQYFFQQWKLKAYANDNHIEIVGDMPIYVABDSDMWANPHLFKTDVN 239

Qy  240 GNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRG 299
          |  | :||  |  :||||||||||:| |  :|| ||||| :|:| | | |:||||||||
Db  240 GKATCIAGCPPDEFSVTGQLWGNPIYDWEAMDKDGYKWWIERLRESFKIYDIVRIDHFRG 299

Qy  300 FEAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFG 359
          ||:||| | |   |.|| ||| || |||||| :  | ||||||||| : :|| ||| ||
Db  300 FESYWEIPAGSDTAAPGEWVKGPGYKLFAAVKEELGELNIIAEDLGFMTDEVIELRERTG 359

Qy  360 FPGMKILQFAFSGEDNAF-LPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAY 418
          |||||||||||  ||:    |:  |  | :  ||.|||| | :|||:|    | : |::|
Db  360 FPGMKILQFAFNPEDESIDSPHLAPA--NSVMYTGTHDNNTVLGWYRN--EIDDA-TREY 414

Qy  419 LARYGIRCLSEYE-VAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWR 477
           :||:  |  :|| | ::  : ||  : ||  : |:|:| | :|||||||| :|| ||:|
Db  415 MARYTNR--KEYETVVHAMLRTVFSSVSFMAIATMQDLLELDEAARMNFPSTLGGNWSWR 472

Qy  478 YAEGDLEPGLAAGLRALAEASQR 500
           :|: :| :|
Db  473 MTEDQLTPAVEEGLLDLTTIYRR 495
```

```
RESULT 8
ABU02622
ID   ABU02622 standard; protein; 505 AA.
XX
AC   ABU02622;
XX
DT   23-OCT-2003  (revised)
DT   11-FEB-2003  (first entry)
XX
DE   S. pneumoniae type 4 strain protein from coding region #2201.
XX
KW   Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
KW   antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
KW   gene therapy; vaccine.
XX
OS   Streptococcus pneumoniae; type 4 strain.
XX
PN   WO20027021-A2.
XX
PD   03-OCT-2002.
XX
PF   27-MAR-2002; 2002WO-IB002163.
XX
PR   27-MAR-2001; 2001GB-00007658.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Masignani V, Tettelin H, Fraser C;
XX
DR   WPI; 2003-040579/03.
DR   N-PSDB; ABX07913.
XX
PT   New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT   useful as medicaments for treating or preventing a disease or infection
PT   due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
PT   ear infection.
XX
PS   Claim 1; SEQ ID NO 4402; 56pp; English.
XX
CC   The invention relates to a protein comprising or having at least 50%
CC   identity to any of the 2469 amino acid sequences, identified in the
CC   specification (available on a computer readable format), or its fragment,
CC   expressed from 2469 identified DNA coding regions from the
CC   Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC   ABS56454. Also included are an antibody which binds one of the proteins,
CC   treating a patient by administering the protein, DNA or antibody (in a
CC   composition), a kit comprising first and second primers, which are the
CC   nucleic acid cited above or fragments between nucleotides 8-100 of a
CC   nucleic acid not defined in the specification, for amplifying a target
```

```
CC      sequence contained within a Streptococcus nucleic acid sequence, where
CC      the first primer is substantially complementary to the target sequence
CC      and the second primer is substantially complementary to the complement of
CC      the target sequence, and where the parts of the primers having
CC      substantial complementarity define the termini of the target sequence to
CC      be amplified, assay comprising contacting a test compound with the
CC      protein, and determining whether the test compound binds to the protein
CC      and a Streptococcus pneumoniae bacterium, where one or more genes
CC      encoding the proteins has been rendered inactive. The proteins, nucleic
CC      acid molecules, antibody and compositions are useful as medicaments for
CC      treating or preventing a disease or infection due to streptococcus
CC      bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
CC      media or ear infection. They are also useful in developing vaccines,
CC      diagnostics and antibiotics. The methods are useful for identifying
CC      immunodominant proteins. The present sequence is one of the 2469 proteins
CC      expressed by the identified coding regions from the genomic sequence.
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
CC      standardise OS field)
XX
SQ      Sequence 505 AA;

Query Match             44.0%;  Score 1194;  DB 6;  Length 505;
Best Local Similarity   48.1%;  Pred. No. 4.2e-107;
Matches  242;  Conservative  77;  Mismatches 168;  Indels  16;  Gaps   9;

Qy      4 QRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQSFS 63
            : |::|:|| ||: ::|:::|:|:  :|::|:: |:  :|    ||| |:::||||||||
Db      3 KRQSGVLMHISSLPGAYGIGSFGQSAYDFVDFLVRTKQRYWQILPLGATSYGDSPYQSFS 62

Qy     64 AFAGNPYLVDPEMLIEKGWLEQSEAPP---PYPTQRVDYGWLYQTRWPLLRRAFAGFRAR 120
            ||||  |:|| : |:: ::|:|: |    :|::::::  :: ||  || ||: | :::|
Db     63 AFAGNTHFIDLDIIVEQGLLEASDLEGVDFGSDASEVDYAKIYYARRPLLEKAVKRFFEV 122

Qy    121 ASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEW-SPELRDREPAALARAREE 179
            || ||  :| :  :::|| ||:||| ||| :::|   |||: ||||  | :|   ::|||
Db    123 GDVKD--FEKFAQDNQSWLELFAEYMAIKEYFDNLAWTEWPDADARARKASALESYREQ 179

Qy    180 LAEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEAD 239
            | ::|  :|  |:|:|:|||| : :|:   |:::|| |::|  |::||::||::|::|: :
Db    180 LADKLVYHRVTQYFFQQWLKLKAYANDNHIEIVGDMPIYVAEDSSDMWANPHLFKTDVN 239

Qy    240 GNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRG 299
            ||   :|: |:| | :||:||||||  | |:| || ||  |:|::|::|:|| |||||||
Db    240 GKATCIAGCPPDEFSVTGQLWGNPIYDWEAMDKDGYKWWIERLRESFKIYDIVRIDHFRG 299

Qy    300 FEAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFG 359
            |:|||:|| |  |  :| |||  |||:::| :||: ::||||||||:: :| :||| :|
Db    300 FESYWEIPAGSDTAAPGEWVKGPGYKLFAAVKEELGNIIAEDLGFMTDEVELERERTG 359

Qy    360 FPGMKILQFAFSGEDNAF-LPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAY 418
            |||||||| ||: ::||: | :|  ::||: |:||||| : |||||  : : |:|:::|:
Db    360 FPGMKILQFAFNPEDESIDSPHLAPA--NSVMYTGTHDNNTVLGWYRN--EIDDA-TREY 414

Qy    419 LARYGIRCLSEYE-VAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWR 477
            | |||: ||::|: | |:::  :::|| : :|  |:: | |||:|:|| || || || :|
Db    415 MARYTNR--KEYETVVHAMLRTVFSSVSFMAIATMQDLLELDEAARMNFPSTLGGNWSWR 472

Qy    478 YAEGDLEPGLAAGLRALAEASQR 500
            ||:  |:||| | ||:| :::|:
Db    473 WTEDQLTPAVEEGLLDLTTIYRR 495
```

RESULT 9
ADQ15028

```
ID      ADQ15028 standard; protein; 487 AA.
XX
AC      ADQ15028;
XX
DT      23-SEP-2004 (first entry)
XX
DE      Clostridium butyricum amylo maltase polypeptide.
```

```
XX
KW      Amylo maltase; transgenic; foodstuff; food additive; modifier;
KW      saccharide; (alpha)-1,4-glucan; cooked rice; Japanese confectionery;
KW      snack; noodle; Chinese dumpling; shao-mai; fishery kneading;
KW      refrigerated; baby food; pet; animal feed; drink; food supplement;
KW      cyclic glucan; Clostridium butyricum.
XX
OS      Clostridium butyricum.
XX
PN      JP2004187674-A.
XX
PD      08-JUL-2004.
XX
PF      12-SEP-2003; 2003JP-00322319.
XX
PR      26-NOV-2002; 2002JP-0342966.
XX
PA      (EZAK ) EZAKI GLICO CO LTD.
PA      (NIDE ) NEC CORP.
XX
DR      WPI; 2004-528719/51.
DR      N-PSDB; ADQ15027.
XX
PT      New mutated amylo maltase polypeptide having increased enzyme activity
PT      and reduced hydrolyzing activity, useful for manufacturing foodstuffs
PT      such as Japanese confectionery, noodles, baby foods, and food additives.
XX
PS      Example 9; SEQ ID NO 19; 51pp; Japanese.
XX
CC      The invention relates to a novel amylo maltase polypeptide comprising the
CC      amino acid sequence of a wild-type amylo maltase with a substitution,
CC      addition or deletion at a position, where the amino acid residue
CC      interacts with acarbose of amino acids other than the wild-type amylo
CC      maltase amino acid sequence. The invention further comprises: a nucleic
CC      acid molecule containing a nucleic acid sequence encoding the novel amylo
CC      maltase polypeptide; a vector containing the nucleic acid sequence; a
CC      cell containing the nucleic acid sequence; biological tissue containing
CC      the nucleic acid sequence; a transgenic organism containing the nucleic
CC      acid sequence; a foodstuff, a food additive or a modifier of foodstuff
CC      containing the nucleic acid sequence; a computer readable recording
CC      medium which contains the information of the nucleic acid sequence
CC      encoding the amino acid sequence of the novel amylo maltase polypeptide;
CC      and a cyclic glucan obtained by reacting the novel amylo maltase
CC      polypeptide on saccharides which have a linear structure of (alpha)-1,4-
CC      glucan.The novel amylo maltase polypeptide is useful for manufacturing
CC      foodstuffs such as cooked rice, Japanese confectionery, snacks, bakeries,
CC      noodles, Chinese dumpling, shao-mai, fishery kneading goods,
CC      refrigerating foods, baby foods, pet foods, feed for animals, drinks, and
CC      food supplements, where the method involves adding the novel amylo
CC      maltase polypeptide to the foodstuff before or immediately after heat
CC      processing of the raw material, where the novel amylo maltase polypeptide
CC      generates a cyclic glucan from the starch of the foodstuff. The novel
CC      amylo maltase polypeptide is useful for manufacturing glucan which has a
CC      cyclic structure of (alpha)-1,4-glycoside linkage, foodstuff materials
CC      and food additives, modifier of foodstuffs, a drink or eatable
CC      composition, infusion solution or composition for adhesion, where the
CC      method involves reacting the saccharide containing the linear structure
CC      of (alpha)-1,4-glycoside linkage or its derivatives with the novel amylo
CC      maltase polypeptide. This sequence represents a Clostridium butyricum
CC      amylo maltase polypeptide of the invention.
XX
SQ      Sequence 487 AA;

Query Match             40.8%;  Score 1108.5;  DB 8;  Length 487;
Best Local Similarity   44.9%;  Pred. No. 8.9e-99;
Matches  215;  Conservative  84;  Mismatches 151;  Indels  29;  Gaps   9;

Qy     11 LHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQSFSAFAGNPY 70
            :|::|| ||:::::|:|| ||       ::   || :|||||||||||||||||:|||||
Db      1 MHISSLPGKYGIGTFGRSAYEFCDFLEKAGQKTWQILPLGQTSYGDSPYQSFSAFAGNPY 60

Qy     71 LVDPEMLIEKGWLEQSEAPPPYP------TQRVDYGWLIYQTRWPLLRRAFAGFRARASA 123
            :|::|| |||||||::|:|: |            :: : |:|  | | || |:::| |: ::
Db     71 TQRVDYGWLYQTRWPLLRRAFAGFRARASA
```

CC   recombinant nucleic acid. The invention also provides a polypeptide or
CC   its fragment essentially having alpha-glucanotransferase activity and
CC   with or without hydrolysing activity. By providing a nucleic acid
CC   encoding such a transferase with a mutation leading to an alteration or
CC   loss of a codon originally encoding a hydrophobic amino acid located in
CC   or around an acceptor, a donor or a catalytic site extending from a
CC   (alphalpha/betabeta) barrel structure of the transferase. The invention
CC   also relates to a bakery ingredient comprising the polypeptide with alpha
CC   -glucanotransferase activity and a bakery product such as bread,
CC   comprising the polypeptide. The transferase is encoded by the recombinant
CC   nucleic acid comprises amylomaltase, a branching enzyme or thermostable
CC   transferase. The transferase is derived from a thermophilic microorganism
CC   e.g. Thermus thermophilus, T.aquaticus or Aquifex aeolicus. The
CC   transferase with enzyme commission (EC) number 2.4.1.25 or 2.4.1.18 is
CC   useful in reducing retrogradation of starch and amylopectin, preferably
CC   in reducing long-term retrogradation of amylopectin. This transferase is
CC   also useful for hydrolysing starch. This sequence represents an Aquifex
CC   aeolicus MTase protein of the invention.
XX
SQ   Sequence 485 AA;

Query Match   40.5%; Score 1100; DB 7; Length 485;
Best Local Similarity   45.8%; Pred. No. 6e-98;
Matches 219; Conservative 74; Mismatches 169; Indels 16; Gaps   6;

Qy   5 RAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWQVLPLGPTSY--GDSPYQSF   62
           |||||||||| |:||| ||| |||||||||||||| ||||| |||||| :|||||
Db   2 RLAGILLHVTSLPSPYGIGDLGKEAYRFLDFLKECGFSLWQVLPLNPTSLEAGNSPYSSN   61

Qy   63 SAFAGNPYLVDPEMLIEKGWLEQSEAPPYPTQRVDYGWLYQTRWPLLRRAFAGFRARAS 122
           | ||||| ::|||||||||||:| | ||:||||||||| ||||| ||||| : :
Db   62 SLFAGNYVLIDPEELLEEDLIKERDL-KRFPLGEALYEVVYKEKELLEKAPFKNFRR--- 117

Qy   123 AQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAREELAE 182
           :: |||: :||| :||||| ||||||||| |||||:|| ||:||:||||||||
Db   118 ---FELLEDFLKEHSYWLRDYALYMAIKEE-EGKEWYEWDEELKRREKEALKRVLNKLKG 173

Qy   183 EVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADGNP 242
           ||||||||:|:||:||||||||||||| |||||:||||||||||||| ||:|||||:|
Db   174 RPYFHVPVFVQFVFKQWEKLRRYARERGISIVGDLPMYPSYSSADVWTNPELFKLDGDLKP 233

Qy   243 TVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGFEA 302
           |||||||||||||||:|||||:||||| :|:||| | | || :|| |||||||||||
Db   234 LFVAGVPDFFSKTGQLWGNPVYNWEHEKEGFRWWIRRVHHNLKLFDFLRLDHFRGFEA   293

Qy   303 YWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGFPG 362
           ||||| |:| || |||||||| || | | || ||||||||||| :|:| |: ||||
Db   294 YWEVPYGEETAVNGRWVKAPGKTLFKKLLSYFPKNPFIAEDLGFITDEVRYLRETFKIPG 353

Qy   363 MKILQFAPSGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRAYLARY 422
           ||||| |||||||| | | :||| || ||| ||||
Db   354 SRVIEFAFYDKESEHLPHN--VEENNVYYTSTHDLPPIRGWFENLGESSRKRLFEYLGRE 411

Qy   423 GIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGPEARMNFPGRLGDNWAWRYAE 480
           | ||:|||| |||||| |||: :|| ||| || | |||| | :|:||: ||
Db   412 ---IKEEKVNEBLIRLVLISRAKFAIIQMQDLLNLGNEARMNYPGRPFGNWRWRIKE 465

RESULT 10
ADC23309
ID   ADC23309 standard; protein; 485 AA.
XX
AC   ADC23309;
XX
DT   18-DEC-2003 (first entry)
XX
DE   Protein of Aquifex aeolicus MTase.
XX
KW   recombinant; alpha-glucanotransferase activity; hydrolysing; hydrolytic;
KW   alphalpha/betabeta; barrel structure; transferase; bakery product; bread;
KW   amylomaltase; branching enzyme; thermostable transferase;
KW   thermophilic microorganism; EC 2.4.1.25; EC 2.4.1.18;
KW   starch retrogradation; amylopectin; alpha-amylase; MTase.
XX
OS   Aquifex aeolicus.
XX
PN   WO2003002728-A2.
XX
PD   09-JAN-2003.
XX
PF   01-JUL-2002; 2002WO-NL000427.
XX
PR   29-JUN-2001; 2001NL-01018426.
XX
PA   (NEDE) NEDERLANDSE ORG TOEGEPAST.
XX
PI   Van Der Maarel MJEC, Dijkhuizen L, Binnema DJ, Van Der Veen BA;
PI   Vos A;
XX
DR   WPI; 2003-201501/19.
DR   N-PSDB; ADC23308.
XX
PT   Novel isolated or recombinant nucleic acid encoding a polypeptide with
PT   alpha glucanotransferase activity, useful for reducing the retrogradation
PT   of starch, amylopectin, and for hydrolyzing starch.
XX
PS   Disclosure; Page 81; 92pp; English.
XX
CC   The invention relates to a novel isolated or recombinant nucleic acid
CC   derived from a nucleic acid encoding a polypeptide essentially having
CC   alpha-glucanotransferase activity but having essentially no hydrolysing
CC   activity, and the recombinant nucleic acid encoding a polypeptide with
CC   hydrolytic activity. The invention further comprises a vector comprising
CC   the recombinant nucleic acid; a host cell comprising the vector or the

Qy   61 FIDLDILNEKNLLDKDD----YEEKNFGDNKEMINYGLIFNEMKVLRKAYMNF----NS 112
           |: | | | | :|| ||||:|:| |||||:|||||||:|
Db   124 QDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAREELAE 183

Qy   113 KDDESAFKFIEDEKKWLDDYSLFMALKYKFNPISWNSWNKDIKLRKNEEIEKYKDELKED 172
           | : | | : :|:||:||:||||| |||| ||| | :||| || || |:||: |
Db   184 VALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAFDSSDVWANPQYFYLEADG-NP 242

Qy   173 VNYWKFLQYEFSQWKNLKDYANKKNIKIIGDIPIYYIAQDSSDVWSNPDIFLLNKETLEP 232
           |:|||| ||:| |:|| || :|||||:|| | ||||||||:|| :| :|: | |
Db   243 TVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHFRGFEA 302

Qy   233 LKVSGCPPDAFSETGQLWGNPIYDWGYLEKTNFEWWVDRIKSSLKLYDILRIDHFRGFEA 292
            | | | | ||||||||||||:|:| |||:|||| |:|||| ||||| :|:|||||||||
Db   303 YWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFGFPG 362

Qy   293 YWSVDYGEKTAQNGKWIKGPEMKLFNVIKEKIGDIEIIAEDLGYLTEETLEFKKRTGFPG 352
           || | |:| ||||:| || |||:|:|:|||:| ||||||| | | ::::|||
Db   353 LKVSGCPPDAFSETGQLWGNPIYDWGYLEKTNFEWWVDRIKSSLKLYDILRIDHFRGFEA 418

Qy   363 MKILQFAFSGE-DNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFR---TAPEAERAFMRAY 418
           ||||||| ||| ||||| ::||||||||||||||||||| ||||:::||
Db   419 LARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGPEARMNFPGRLGDNWAWR   477

Qy   353 MKIIQFAFGGDSSNPYLPHNY--EKNCVAYTGTHDNDTVRGWFEVTGSKEEKEKA----- 405
           ||:||||| | |:|||||||| ||:||||||| |||| :||| |||||
Db   406 -VEY-FKLTEEEGYNWGVIRGVWSSVANTSIGVMQDFLNLGNEARINKPSTLASNWSWR 462

RESULT 11
ADR94304
ID   ADR94304 standard; protein; 458 AA.
XX
AC   ADR94304;
XX
DT   16-DEC-2004 (first entry)
XX
DE   Novel S. pneumoniae protein sequence, SEQ ID 2939.
XX
KW   Meningitis; bacteraemia; pneumonia; otitis media; vaccine;
KW   bacterial infection.
XX
OS   Streptococcus pneumoniae.
XX

```
Db         :||  ||: ||||||||||:|| |  ||:|:||||||:| |||:
       299 FMTDSVIELRERTGFPGMKILQFAFNPEDESIDSPHLAPA--NSVMYTGTHDNNTVLGWY 356

Qy     405 RTAPEAERAFMRAYLARYGIRCLSEYE-VAGALIELAFKSPAKLAIVPLQDVLGLGPEAR 463
           |||   | | |:||| |:|  |:    |||  | : ::|| |
Db     357 RN--EIDDA-TREYMARYTNR--KEYETVVHAMLRTVFSSVSFMAIATMQDLLELDEAAR 411

Qy     464 MNFPGRLGDNWAWRYAEGDLEPGLAAGLRALAEAASQR 500
           ||||| :|||| | ||:|:|| |
Db     412 MNFPSTLGGNWSWRMTEDQLTPAVEBGLLDLTTIYRR 448

RESULT 12
AEA58174
ID   AEA58174 standard; protein; 458 AA.
XX
AC   AEA58174;
XX
DT   25-AUG-2005  (first entry)
XX
DE   Streptococcus pneumoniae ORF amino acid sequence SEQ ID NO:2939.
XX
KW   bacterial infection; Streptococcus pneumoniae infection; antibacterial;
KW   vaccine.
XX
OS   Streptococcus pneumoniae.
XX
PN   US2005136404-A1.
XX
PD   23-JUN-2005.
XX
PF   10-JUL-2003; 2003US-00617320.
XX
PR   02-JUL-1997;  97US-0051553P.
PR   12-MAY-1998;  98US-0085131P.
PR   30-JUN-1998;  98US-00107433.
XX
PA   (DOUC/) DOUCETTE-STAMM L A.
PA   (BUSH/) BUSH D.
XX
PI   Doucette-Stamm LA,    Bush D;
XX
DR   WPI; 2005-477576/48.
DR   N-PSDB; AEA55571.
XX
PT   New isolated nucleic acid molecules and encoded polypeptides useful for
PT   diagnosing, preventing or treating bacterial infections, particularly
PT   Streptococcus pneumoniae infection.
XX
PS   Claim 5; SEQ ID NO 2939; 144pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule for detecting,
CC   preventing or treating pathological conditions resulting from bacterial
CC   infection. The isolated nucleic acid comprises: (a) any of the 2603
CC   nucleotide sequences of AEA55236 to AEA57838; (b) a nucleotide sequence
CC   encoding a Streptococcus pneumoniae polypeptide comprising any of the
CC   2603 amino acid sequences of AEA57839 to AEA60441; or (c) a nucleotide
CC   sequence of at least 8 nucleotides in length, where the sequence is
CC   hybridizable to a nucleic acid having any of the nucleotide sequences in
CC   (a). Also described: (1) a recombinant expression vector comprising the
CC   above nucleic acid operably linked to a transcription regulatory element;
CC   (2) a cell comprising the recombinant expression vector; (3) producing an
CC   S. pneumoniae polypeptide; (4) a probe comprising a nucleotide sequence
CC   consisting of at least 8 nucleotides of any of AEA55236 to AEA57838; (5)
CC   treating a subject for S. pneumoniae infection; (6) a recombinant or
CC   substantially pure preparation of an S. pneumoniae polypeptide or its
CC   fragment, where the polypeptide is selected from AEA57839 to AEA60441;
CC   (7) a vaccine composition for preventing or treating an S. pneumoniae
CC   infection, comprising an amount of the above nucleic acid or polypeptide;
CC   (8) detecting the presence of a Streptococcus nucleic acid in a sample;
CC   (9) a computer readable medium having recorded the nucleotide sequences
CC   of AEA55236 to AEA57838; (10) a computer based system for identifying
CC   fragments of the Streptococcus genome of commercial importance. The
```

```
PN   US6800744-B1.
XX
PD   05-OCT-2004.
XX
PF   30-JUN-1998;   98US-00107433.
XX
PR   02-JUL-1997;   97US-0051553P.
PR   12-MAY-1998;   98US-0085131P.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Doucette-Stamm LA,   Bush D;
XX
DR   WPI; 2004-697205/68.
DR   N-PSDB; ADR91701.
XX
PT   New isolated nucleic acid encoding a Streptococcus pneumoniae
PT   polypeptide, useful for diagnosing, preventing and/or treating
PT   pathological conditions resulting from the bacterial infection.
XX
PS   Disclosure; SEQ ID NO 2939; 151pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising a sequence
CC   encoding a Streptococcus pneumoniae ADR9136polypeptide, or its
CC   fragments, with any of 9 fully defined sequences (appearing as ADR94308,
CC   ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,
CC   ADR96079) or any of the fully defined sequences appearing as ADR91705,
CC   ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or
CC   ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide
CC   sequences, or at least 40, 60 or 300 consecutive nucleotides, which is
CC   hybridisable under high stringency conditions to the nucleotide sequence.
CC   The nucleic acids and proteins are chosen from 5206 disclosed sequences.
CC   Also included are a recombinant expression vector comprising the isolated
CC   nucleic acid cited above operably linked to a transcription regulatory
CC   element, a cell comprising the recombinant expression vector and a probe
CC   comprising at least 20 consecutive nucleotides of the nucleotide
CC   sequences as cited above. The methods and compositions of the present
CC   invention are useful for the diagnosis, prevention and/or treatment of
CC   pathological conditions resulting from bacterial infection by
CC   Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and
CC   otitis media. The present sequence is one of the 2603 disclosed S.
CC   pneumoniae protein sequences. Note:The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=6800744B1.
XX
SQ   Sequence 458 AA;

Query Match           39.6%;    Score 1076;  DB 8;   Length 458;
Best Local Similarity 48.4%;    Pred No. 1.2e-95;
Matches 221;   Conservative  66;  Mismatches 154;   Indels  16;   Gaps    9;

Qy     50 GPTSYGDSPYQSFSAFAGNPYLVDPEMLIEKGWLEQSEAPP--PYPTQRVDYGWLYQTR 106
          ||| |||||||||||||||||||     |: |:|:|| ||| |        :|  ||:
Db      2 GATSYGDSPYQSFSAFAGNTHFIDLDILVEQGLLEASDLEGVDFGSDASEVDYAKIYYAR  61

Qy    107 WPLLRRAFAGFRARASAQDKTRLEAFIEAAERFWLEDYALFMALKTRFDGKPWNEW-SPEL 165
          |||   || |  | |     |:|||||| ||:|||| |||| |||| ||  ||||| |||
Db     62 RPLLEKAVKRFFEVGDVKD--FEKFAQDNQSWLELFAEYMAIKEHFDNLAWTEWPDADA 118

Qy    166 RDREPAALARAREELAEEVALYEWIQWLFYLEWGGTKAYAESKGIQIIGDMPIFVAFDSS 225
          |:|::|| |   |:|: | ||    :||     :|| ||  :| | : |||||||||||
Db    119 RARKASALESYREQLADKLVYHRVTQYFFFQQWLKLKAYANDNHIEIVGDMPIYVAEDSS 178

Qy    226 DWANPQVFYLEADGNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQS 285
          |||||  ||| |||||||||  | | |||  |||||| |||:|  ||:||| |: |::
Db    179 DMWANPHLFKTDVNGKATCIAGCPPDEFSVTGQLWGNPIYDWEAMDKDGYKWWIERLRES 238

Qy    286 LKQCHLVRIDHFRGFEAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLG 345
          |  |:||  |||||| |  |  |  | | | ||  :|| ||||||||:|:||||:||| |
Db    239 FKIYDIVRIDHFRGFESYWEIPAGSDTAAPGEWVKGPGYKLFAAVKEELGELNIIAEDIG 298

Qy    346 VITPEVEALRDGFGFPGMKILQFAFSGEDNAF-LPHNYPAHGNVVVYSGTHDNDTTLGWF 404
```

```
CC   composition and methods are useful for diagnosing, preventing or treating
CC   bacterial infections, particularly S. pneumoniae infection. The present
CC   sequence represents a S. pneumoniae ORF amino acid sequence from the
CC   present invention. Note - The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from the USPTO web site.
XX
SQ   Sequence 458 AA;

Query Match             39.6%;  Score 1076;  DB 9;  Length 458;
Best Local Similarity   48.4%;  Pred. No. 1.2e-95;
Matches  221;  Conservative  66;  Mismatches 154;  Indels 16;  Gaps  9;

Qy    50  GPTSYGDSPYQSFSAFAGNPYIVDPEMLIEKGWLEQSEAPP--PYPTQRVDYGWLYQTR  106
          | |||||||||||| ||    :.|  :.| :||||  : :|      :.|
Db     2  GATSYGDSPYQSFSAFAGNTHFIDLDILVEGLLEASDLEGVDFGSDASEVDYAKIYYAR   61

Qy   107  WPLLRRAFAGPFRARASAQDKTRLEAFIEAERFWLEDYALFMALKTRFDGKPWNEW-SPEL  165
          .|||..| .|||    : :      | : ||:|||| |: :||||:|     .|  | .
Db    62  RPLLEKAVKRPFEVGDVKD---FEKFAQDNOSWLELFAEYMAIKEHFDNLAWTEWPDADA  118

Qy   166  RDREPAALARAREELAEEVALYEWIQMLFVLEWGQTKAYAESKGIQIIGDMPIFVAFDSS  225
          |   .: |   : : .  :: .  : | : ::| |  : | |.  |:|:||||:|| |||
Db   119  RARKASALESYREQLADKLVTHRVTQYFFQQWLKLKAYANDNHIEIVGDMPIYVAEDSS  178

Qy   226  DWWANPQYFYLEADGNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQS  285
          |:WANP: | . | .     |. | ||:|:||||WGNP:| | : : | .:| | | |.|
Db   179  DMWANPHLFKTDVNGKATCIAGCPPDEFSVTGQLWGNPIYDWEAMDKDGYKWWIERLRES  238

Qy   286  LKQCHLVRIDHFRGFEAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLG  345
          .|  .:VRIDHFRGFE:YWE:P.G   A  |.WVK.PG KLFAAV:  | . |:IAEDLG
Db   239  FKIYDIVRIDHFRGFESYWEIPAGSDTAAPGEWVKGPGYKLFAAVKEELGELNIIAEDLG  298

Qy   346  VITPEVEALRDGFGPPGMKILQFAFSGEDNAF-LPHNYPAHGNVVVYSGTHDNDTTLGWF  404
          .:| :: .:: . |PGMKILQFAF. ||:. . .| :  | |   |:GTHDN:T.LGW:
Db   299  FMTDBVIELERRTGFPGMKILQFAFNPEDESIDSPHLAPA--NSVMYTGTHDNNTVLGWY  356

Qy   405  RTAPEAERAFMRAYLARYGIRCLSEYE-VAGALIELAFKSPAKLAIVPLQDVLGLGPEAR  463
          |     : :  . | |||  :  .:.  : . .:. |::   .||:| | |:| |  |.|
Db   357  RN--EIDDA-TREYMARYTNR--KEYETVVHAMLRTVFSSVSFMAIATMQDLLELDEAAR  411

Qy   464  MNFPGRLGDNWAWRYAEGDLEPGLAAGLRALAEASOR  500
          MNFP. LG.NW:WR  | | |.   | || ..    |
Db   412  MNFPSTLGGNWSWRMTEDQLTPAVEEGLLDLTTIYRR  448


RESULT 13
ABP28631
ID   ABP28631 standard; protein; 497 AA.
XX
AC   ABP28631;
XX
DT   02-JUL-2002  (first entry)
XX
DE   Streptococcus polypeptide SEQ ID NO 6438.
XX
KW   Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW   group A streptococcus; Streptococcus pyogenes; antibacterial;
KW   antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
OS   Streptococcus pyogenes.
XX
PN   WO200234771-A2.
XX
PD   02-MAY-2002.
XX
PF   29-OCT-2001; 2001WO-GB004789.
XX
PR   27-OCT-2000; 2000GB-00026333.
PR   24-NOV-2000; 2000GB-00028727.
PR   07-MAR-2001; 2001GB-0005640.
XX
PA   (CHIR-) CHIRON SPA.


PA   (GENO-) INST GENOMIC RES.
XX
PI   Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;
PI   Tettelin H;
XX
DR   WPI; 2002-352536/38.
DR   N-PSDB; ABN69262.
XX
PT   New Streptococcus protein for the treatment or prevention of infection or
PT   disease caused by Streptococcus bacteria, such as meningitis, and for
PT   detecting a compound that binds to the protein.
XX
PS   Claim 1; Page 3806; 4525pp; English.
XX
CC   The invention relates to a protein (ABP25413-ABP30895) from group B
CC   streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC   (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
CC   the specification. The proteins have antibacterial and antiinflammatory
CC   activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC   antibodies that bind (I) are used in the manufacture of medicaments for
CC   the treatment or prevention of infection or disease caused by
CC   Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC   Nucleic acids encoding (I) are used to detect Streptococcus in a
CC   biological sample. (I) is used to determine whether a compound binds to
CC   (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC   used as a vaccine or diagnostic composition. The disease caused by
CC   Streptococcus that is prevented or treated may be meningitis. Nucleic
CC   acid encoding (I) may be used to recombinantly produce (I) and may be
CC   used in gene therapy. Antibodies to (I) are used for affinity
CC   chromatography, immunoassays, and distinguishing/identifying
CC   Streptococcus proteins
XX
SQ   Sequence 497 AA;

Query Match             38.4%;  Score 1041;  DB 5;  Length 497;
Best Local Similarity   43.6%;  Pred. No. 3.6e-92;
Matches  212;  Conservative  81;  Mismatches  81;  Indels 16;  Gaps  8;

Qy     4  QRAFGILLHPTSPPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPTSYGDSPYQSFS   63
          :|| |||:|  | ||::|||. |:.|:||.: ||| |:  |:|:||| :|:|||||||||
Db     3  KRASGILMHISSLPGKFGIGTFGKSAFEFVDFLAETKQTYWQILPLTTTSFGDSPYQSFS   62

Qy    64  AFAGNPYLVDPEMLIEKGWLEQSE-APPPYPT--QRVDYGWLYQTRWPLLRRAFAGPFRAR  120
          |:AGN :::D ||::::  :|  : |    .|   .VDY. |:|.|.PLL::|. . :.
Db    63  AIAGNTHFIDFELLVDDELLEAADLCDITFGTNPEAVDYAQLFQVRRPLLEKAVRAF--V  120

Qy   121  ASAQDKTRLAFIEAERFWLEDYALFMALKTRFDGKPWNEWSPE-LRDREPAALARAREE  179
          |.|:: :.|.|.|   :|  .:|.|FMALK  ::..| ::  :     :  |  ..|:.|
Db   121  AEQENVCKLEAF-ETASSWLTDFAEFMALKEYFNNKALQDWDDETVIKRQEDSLNNYREL  179

Qy   180  LAEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAPDSSDVWANPQYFYLEAD  239
          ||::::.|..: :::| : .  :| |:. |||:IIGDMPI:V:.|||:VW::P: |: .|
Db   180  LAKKITYHKVCQYFYQQWSALKTYANHKGIEIIGDMPIYVSADSVEVWTMPELFKVDSD  239

Qy   240  GNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNPAWCIARIRQSLKQCHLVRIDHFRG  299
          .:P :.||P.| |||:|GQLWGNP Y.W..:|:.NPAW.I.||:||:|.  |.RIDHF:G
Db   240  KKPLFIAGVPADGFSEDGQLWGNPTYNWSAHEKSNPAWWIYRIQESFKLYDQLRIDHFKG  299

Qy   300  FEAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDGFG  359
          |. .WE:P.G   :   |W:.APG  LF:AVR  |.. .IIAE:LG :  |. :.: ..G
Db   300  FSDFWEIPAGDKTARNGHWASAPGIALFSAVREALGELPIIAENLGYIDEKAEQLLASTG  359

Qy   360  FPGMKILQFAFSG--EDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMRA  417
          FPGMKIL:F.: .   :. :LPH :.:   .. |:GTHDN::..GW: : | ::. .. .
Db   360  FPGMKILEFGLFDITSQSIDLPHYDR--NCVAYTGTHDNEVVNGWYDNLSEEQVHFVNN  417

Qy   418  YLARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAWR  477
          YL :    .:.:   ::::  . .|:.. . :  .:|:|     :RMN:P  .|.NW.WR
Db   418  YLHKH----ADESITKAMLRTIFASVCDTAILCIQDLLDKDGKSRMNMPNTIGGNWQWR  472

Qy   478  YAEGDL  483
          :.:|:|
Db   473  MLDGEL  478
```

RESULT 14
ADQ15022
ID   ADQ15022 standard; protein; 576 AA.
XX
AC   ADQ15022;
XX
DT   23-SEP-2004 (first entry)
XX
DE   Potato amylo maltase polypeptide.
XX
KW   Amylo maltase; transgenic; foodstuff; food additive; modifier;
KW   saccharide; (alpha)-1,4-glucan; cooked rice; Japanese confectionery;
KW   snack; noodle; Chinese dumpling; shao-mai; fishery kneading;
KW   refrigerated; baby food; pet; animal feed; drink; food supplement;
KW   cyclic glucan; potato.
XX
OS   Solanum tuberosum.
XX
PN   JP2004187674-A.
XX
PD   08-JUL-2004.
XX
PF   12-SEP-2003; 2003JP-00322319.
XX
PR   26-NOV-2002; 2002JP-00342966.
XX
PA   (EZAK) EZAKI GLICO CO LTD.
PA   (NIDE) NEC CORP.
XX
DR   WPI; 2004-528719/51.
DR   N-PSDB; ADQ15021.
XX
PT   New mutated amylo maltase polypeptide having increased enzyme activity
PT   and reduced hydrolyzing activity, useful for manufacturing foodstuffs
PT   such as Japanese confectionery, noodles, baby foods, and food additives.
XX
PS   Example 9; SEQ ID NO 13; 51pp; Japanese.
XX
CC   The invention relates to a novel amylo maltase polypeptide comprising the
CC   amino acid sequence of a wild-type amylo maltase with a substitution,
CC   addition or deletion at a position, where the amino acid residue
CC   interacts with acarbose of amino acids other than the wild-type amylo
CC   maltase amino acid sequence. The invention further comprises: a nucleic
CC   acid molecule containing a nucleic acid sequence encoding the novel amylo
CC   maltase polypeptide; a vector containing the nucleic acid sequence; a
CC   cell containing the nucleic acid sequence; biological tissue containing
CC   the nucleic acid sequence; a transgenic organism containing the nucleic
CC   acid sequence; a foodstuff, a food additive or a modifier of foodstuff
CC   containing the nucleic acid sequence; a computer readable recording
CC   medium which contains the information of the nucleic acid sequence
CC   encoding the amino acid sequence of the novel amylo maltase polypeptide;
CC   and a cyclic glucan obtained by reacting the novel amylo maltase
CC   polypeptide on saccharides which have a linear structure of (alpha)-1,4-
CC   glucan. The novel amylo maltase polypeptide is useful for manufacturing
CC   foodstuffs such as cooked rice, Japanese confectionery, snacks, bakeries,
CC   noodles, Chinese dumpling, shao-mai, fishery kneading goods,
CC   refrigerating foods, baby foods, pet foods, feed for animals, drinks, and
CC   food supplements, where the method involves adding the novel amylo
CC   maltase polypeptide to the foodstuff before or immediately after heat
CC   processing of the raw material, where the novel amylo maltase polypeptide
CC   generates a cyclic glucan from the starch of the foodstuff. The novel
CC   amylo maltase polypeptide is useful for manufacturing glucan which has a
CC   cyclic structure of (alpha)-1,4-glycoside linkage, foodstuff materials
CC   and food additives, modifier of foodstuffs, a drink or eatable
CC   composition, infusion solution or composition for adhesion, where the
CC   method involves reacting the saccharide containing the linear structure
CC   of (alpha)-1,4-glycoside linkage or its derivatives with the novel amylo
CC   maltase polypeptide. This sequence represents a potato amylo maltase
CC   polypeptide of the invention.
XX
SQ   Sequence 576 AA;

Query Match   38.2%; Score 1035.5; DB 8; Length 576;
Best Local Similarity   44.9%; Pred. No. 1.5e-91;
Matches 216; Conservative 72; Mismatches 178; Indels 15; Gaps   6;

Qy   4 QRAFGILLHPTSFPGRWGIGALGREAERFLDWLADAGARWWQVLPLGPT----SYGDSPY 59
              :| |||||||||||||| :   : ||||| :|||||| : || || :||
Db   77 RRRAGILLHPTSFPGPYGIGDLGPQAFKFLDWLHLAGCSLWQVLPLVPPGKRGNEDGSPY 136

Qy   60 QSFSAFAGNPYLVDPEMLIEKGWLEQSEAPPYPTQRVDYGWLYQTRWPLLRRAFAGFRA 119
              |||| ||| | :||: |:| : | | :|||::| ||: ||: | : | |::| :|
Db   137 SGQDANCGNTLLISLELVDDGLLKMEELPEPLPTDRVNYSTISEIKDPLITKA-AKRLL 195

Qy   120 RASAQDKTRLEAFIRAERF--WLEDYALFMALKTRFDGKPWNEWSPELRDREPAALARAR 177
              ||:| : : :| : | :|| | : | | : | | :|| ||:|
Db   196 SSEGELKDQLENFRRDPNISSWLEDAYFAAIDNSVNTISWYDWPEPLKNRHLAALEEVY 255

Qy   178 EELAEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAPDSSDWANPQYFYLE 237
              : | : | | ||:||:||||||| | ||||:||||||:| ||:   ::|::| :|
Db   256 QSEKDFDIFIAQQFLFQRQWKKVRDYARSKGISIMGDMPIYVGYHSADVWANKKQFLLN 315

Qy   238 ADGNPTVVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRIDHF 297
              | | | |||| ||||||||||| ||||| : : : || | : | |||||
Db   316 RKGFPLIVSGVPPDAFSETGQLWGSPLYDWKAMEKDGFSWWVRRIQRATDLPDEFRIDHF 375

Qy   298 RGFEAYWEVPFGRPNAVEGRWVKAPGBKLFAAVRAQLSDAPIIAEDLGVITPEVEALRDG 357
              ||| | | |||| | |||:||| ||| | |:|||||||||||||| | ||:||
Db   376 RGFAGFWAVPSBEKIAILGRWKVGPGKPLFDAILQVGKINIIAEDLGVITBDVVQLRKS 435

Qy   358 FGFPGMKILQFAF-SGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAERAFMR 416
              |||||||||| | ||:: : | || | : :||||||:|||| |||| |:||| |
Db   436 IEAPGMAVLQFAFGSDAENPHLPHNH--EQNQVVTGTHDNDTIRGWWDTLPQEEKSNVL 493

Qy   417 AYLARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGDNWAW 476
              ||:| | : |:|| || |||||| || ||||: :||||||||:||||||||
Db   494 KYLSN-----IEEEISRGLIEGAVSSVARIAIIPMQDVLGLGSDSRMNIPATQFGNWSW 548

Qy   477 R 477
              |
Db   549 R 549

RESULT 15
ABP28630
ID   ABP28630 standard; protein; 498 AA.
XX
AC   ABP28630;
XX
DT   02-JUL-2002 (first entry)
XX
DE   Streptococcus polypeptide SEQ ID NO 6436.
XX
KW   Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW   group A streptococcus; Streptococcus pyogenes; antibacterial;
KW   antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
OS   Streptococcus agalactiae.
XX
PN   WO200234771-A2.
XX
PD   02-MAY-2002.
XX
PF   29-OCT-2001; 2001WO-GB004789.
XX
PR   27-OCT-2000; 2000GB-00026333.
PR   24-NOV-2000; 2000GB-00028727.
PR   07-MAR-2001; 2001GB-00005640.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Telford J, Masignani V, Margarit Y RosI, Grandi G, Fraser C;
PI   Tettelin H;

Tue Jan 17 09:28:41 2006

Page 12

us-10-618-976-2.rag

```
XX
DR    WPI; 2002-352536/38.
DR    N-PSDB; ABN69261.
XX
PT    New Streptococcus protein for the treatment or prevention of infection or
PT    disease caused by Streptococcus bacteria, such as meningitis, and for
PT    detecting a compound that binds to the protein.
XX
PS    Claim 1; Page 3805; 4525pp; English.
XX
CC    The invention relates to a protein (ABP25413-ABP30895) from group B
CC    streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC    (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC    the specification. The proteins have antibacterial and antiinflammatory
CC    activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC    antibodies that bind (I) are used in the manufacture of medicaments for
CC    the treatment or prevention of infection or disease caused by
CC    Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC    Nucleic acids encoding (I) are used to detect Streptococcus in a
CC    biological sample. (I) is used to determine whether a compound binds to
CC    (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC    used as a vaccine or diagnostic composition. The disease caused by
CC    Streptococcus that is prevented or treated may be meningitis. Nucleic
CC    acid encoding (I) may be used to recombinantly produce (I) and may be
CC    used in gene therapy. Antibodies to (I) are used for affinity
CC    chromatography, immunoassays, and distinguishing/identifying
CC    Streptococcus proteins
XX
SQ    Sequence 498 AA;

Query Match      37.6%;   Score 1020;  DB 5;  Length 498;
Best Local Similarity   40.7%;   Pred. No. 4e-90;
Matches  207;   Conservative  89;   Mismatches  186;   Indels  26;   Gaps  9;

Qy      4  QRAFGILLHPTSFPPGRWGIGALGREABRFLDWLADAGARWWQVLPLGPTSYGDSPYQSFS   63
           : ||:|:|||:|||:||:||  ||||  :||:|||  :||:||  ||||||||||||||||
Db      3  KRASGVLMHITSLPGDLGIGTFGREAYAFVDFLVETDQKFWQILPLTTTSFGDSPYQSFS   62

Qy     64  AFAGNPYLVDPEMLIEKGWLEQSEAPPYPT------QRVDYGWLYQTRWPLLRRAPAG   116
           :||| |||::||:: |  :::|  || |     :|| :|:::|:: |: :|:
Db     63  AVAGNTHLIDFDLLTLSGFISKDD----YQNISFGQDPEVVDYAGLFEKRRPVLEKAVKN  118

Qy    117  FRARASAQDKTR-LEAFIEABRFWLEDYALFMALKTRFDGKPWNEWSPE-LRDREPAALA   174
           |: :  :||| :||:||  |:|||||:|  |:|:|:|| : | |:|:  |||||  |||
Db    119  FLQEERA---TRMLSDFLQEEK-WVTPFAEFMAIKEHFGNKALQEWDDKAIIRREEEALA  174

Qy    175  RAREELAEEVALYEWIQWLFYLEWGQTKAYAESKGIQIIGDMPIFVAPDSSDVWANPQYF   234
           | : :|:| :| : ||||||||  ::| ||||:|:|||||||||  :|| :|| ||
Db    175  GYRQKLSEVIKYHEVTQYFFYKQWFELKEYANDKGIQIIGDMPIYVSADSVEVWTMPELF  234

Qy    235  YLEADGNPTVAGVPRDYFSETGQLWGNPLYRWDVMERDNFAWCIARIRQSLKQCHLVRI   294
           |:  ||||: :||||| :   |:: ||| |:|:||:  :|| :||: ::|:||:|:
Db    235  KLDRDKQPLAIAGVPADDFSDDGQLWGNPIYNMDYHKESDFDWWIYRIQSGVKMYDYLRI  294

Qy    295  DHFRGFEAYWEVPFGRPNAVEGRWVKAPGEKLFAAVRAQLSDAPIIAEDLGVITPEVEAL   354
           |||| ||  || :| || :|   || :||  ::: ||  :|:|| |||||||::|| |
Db    295  DHFKGFSDYDWEIRGDYQTANDGSWQPAPGPELFATIKEKLGDLPIIAENLGYIDERAERL  354

Qy    355  RDGFGFPGMKILQFAF--SGEDNAFLPHNYPAHGNVVVYSGTHDNDTTLGWFRTAPEAER   412
           |||| ||||||:||    : ||||  |:   ::|||||     :|||  | |
Db    355  LAGTGFPGMKIMEFGFYDTTGNSIDIPHNYTE--NTIAYAGTHDNEVINGWFENLTVEQK  412

Qy    413  AFMRAYLARYGIRCLSEYEVAGALIELAFKSPAKLAIVPLQDVLGLGPEARMNFPGRLGD   472
           :|   :|: | || ::     |  ||||: ||   :||    ||        ||::
Db    413  AYAENYMRR-----LPNEPITETVLRTLYATVSQTTITCMQDLLDKPADSRMNMPNTVGG  467

Qy    473  NWAWRYAEGDLEPGLAAGLRALAEASQR  500
           || |  |||:|   |  |::|  |
Db    468  NWQWRMRKEDLITENRKAFLKEITTIYNR  495
```

Search completed: January 10, 2006, 14:09:43
Job time : 137 secs